

List-decoding thresholds of approximate polymorphisms:
 $s_g^L < s_g^U$, $s_g < s_g^U$, an exact characterisation of s_g , and when s_g is
nontrivial
Paata Ivanisvili

This proof was generated on October 21, 2026, by Grok Bot using Cursor's agents, guided by several prompts from Paata Ivanisvili

On Open Questions 5–6 of Chase–Filmus–Minzer–Mossel–Saurabh

22 October 2026

Abstract

For a Boolean function $g: \{-1, 1\}^m \rightarrow \{-1, 1\}$ with $\mathbb{E}g \neq 0$, Chase, Filmus, Minzer, Mossel and Saurabh (STOC 2022, “Approximate polymorphisms”, henceforth CFMMS) introduced the list-decoding threshold s_g and proved $s_g^L \leq s_g \leq s_g^U$ for two Gaussian optimisation problems that differ only in whether the m Gaussian test functions are required to coincide. They asked whether $s_g^U = s_g^L$ (Open Question 5) and for which g the threshold exceeds the trivial value $\frac{1}{2} + \frac{1}{2}|\mathbb{E}g|$ (Open Question 6). We prove:

- (A) For $g = \text{E2}$ (“exactly two of the three inputs equal +1”), $s_g^L = \frac{5}{8}$ exactly, while $s_g^U > \frac{5}{8}$ and $s_g > \frac{5}{8}$. Hence $s_g^U \neq s_g^L$ and $s_g^L \neq s_g$ in general, and CFMMS Lemma 5.6 (nontriviality of s_g^L when all level-one Fourier coefficients are nonzero) is false. Moreover $0.6251187 < s_g \leq s_g^U < 0.626951$ and $s_g^U > 0.6253166$ (computer-assisted; the upper bound uses Borell’s noise stability theorem, and the CFMMS argument gives 0.65067).
- (B) An exact characterisation $s_g = s_g^{\text{lab}}$, where s_g^{lab} is a *labelled* Gaussian optimisation problem lying between s_g^L and s_g^U : finitely many “label” bits select which Gaussian test function is used, and the joint law of the labels of the m columns and of the output is remembered.
- (C) Nontriviality criteria: with $J = \{j : \hat{g}(\{j\}) \neq 0\}$ and $h(w) = \mathbb{E}_u g(w, u)$ ($w \in \{-1, 1\}^J$), s_g^U is nontrivial iff $\text{sgn}(\mathbb{E}g) h(w) < 0$ for some w , and this condition is necessary for s_g ; it is sufficient for s_g whenever $\|h\|_2^2 > |\mathbb{E}g|$ (in particular whenever all level-one coefficients are nonzero, which repairs CFMMS Lemma 5.6 with s_g in place of s_g^L), and more generally whenever a finite “labelled decoder” certificate exists. For all g on $m \leq 4$ variables (exact computer check of all 52,144 relevant functions), s_g is nontrivial iff s_g^U is nontrivial iff $\text{sgn}(\mathbb{E}g)h(w) < 0$ for some w .
- (D) An exact criterion: $s_g > \text{triv}_g$ iff a finite “labelled decoder” certificate exists (a condition on h only).
- (E) For an explicit g_0 on 5 variables, $s_{g_0}^L = s_{g_0} = \text{triv}_{g_0} = \frac{3}{4} < s_{g_0}^U$. Hence the CFMMS upper bound s_g^U is not tight either, and for every $m \geq 5$ the s_g^U criterion of (C) does not characterise nontriviality of s_g (while it does for $m \leq 4$ and $\mathbb{E}g \neq 0$). The proof rests on a sharp L^1 -Khinchine inequality for q -biased bits, $\mathbb{E}|\sum_i \alpha_i \chi_i| \leq \mathbb{E}|\chi_1|$ for unit α , equivalently $W^1(f) \leq q(1-q)$ for $f: \{-1, 1\}^n \rightarrow [0, 1]$ on the q -biased cube, proved for every $q \in [\frac{3}{8}, \frac{5}{8}]$ (Theorem K; computer-assisted except at $q \in \{\frac{3}{8}, \frac{2}{5}, \frac{5}{12}, \frac{1}{2}\}$ and their images under $q \mapsto 1-q$), false for $q < 1 - \frac{1}{\sqrt{2}}$. A second explicit function g_1 (using $q = \frac{2}{5}$) behaves in the same way, and so does a third, g_2 , whose proof instead uses a Hoffman-type level-one inequality $W^1(f) \leq q\mathbb{E}f^2 + \frac{1-2q}{q}(\mathbb{E}f)^2$ for all $f \geq 0$ on the q -biased cube, $0 < q \leq \frac{1}{2}$ (Theorem L). Together with their symmetric images these settle 48 of the 288 undecided five-variable functions. Coupling arguments, which replace each row of the label matrix by a row of three to seven correlated entries realising the same pair (and triple)

laws, settle 192 more and give second proofs of the cases of g_0 and g_1 that do not use Theorem K; none of the remaining 48 (two problems) has a certificate with at most three label bits (computer-assisted, exact), and each of the two problems is equivalent to an explicit inequality for a single function $P: \{-1, 1\}^k \rightarrow [0, 1]$, which holds numerically for $k \leq 8$.

- (F) A maximal-correlation criterion for triviality when h has no terms of degree ≥ 3 (Theorem F), proved with Witsenhausen's tensorisation of maximal correlation. It gives an explicit g_F on 6 variables with only two relevant coordinates ($|J| = 2$) and $s_{g_F}^L = s_{g_F} = \text{triv}_{g_F} = \frac{9}{16} < s_{g_F}^U$, and such functions on every $m \geq 6$; for $|J| \leq 2$ and $m \leq 5$ none exist (computer-assisted). If the nonzero level-one coefficients of g are pairwise distinct, the s^U criterion of (C) is exact. Two refinements for one class of size two, which feed Theorem L and a second coupling into the covariance bound, decide every g with $|J| \leq 2$ and $m \leq 8$ (computer-assisted, exact): $s_g > \text{triv}_g$ iff a certificate as in (D) exists with at most three rows; two rows suffice for $m \leq 7$ but not for $m = 8$.

Contents

1	Setting and results	3
1.1	Notation and the CFMMS definitions	3
1.2	The labelled value	4
1.3	Results	5
2	Preliminaries	8
3	Theorem A(a): $s_{E2}^L = 5/8$	10
4	Theorem A(b)–(d): bounds for s_{E2}^U and s_{E2}	12
4.1	Lower bounds	12
4.2	An upper bound for s_{E2}^U : proof of Theorem A(d)	13
5	The lower bound $s_g \geq s_g^{\text{lab}}$	18
6	The upper bound $s_g \leq s_g^{\text{lab}}$	20
6.1	Tools	20
6.2	A three-party noise lemma	22
6.3	Proof of Theorem 6.1	23
7	Nontriviality: proof of Theorem C	23
7.1	Small arity: proof of Theorem C(v)	24
8	Labels absorb the Gaussian part: proof of Theorem D	25
9	An explicit g with $s_g = \text{triv}_g < s_g^U$: proof of Theorem E	26
9.1	The function g_0	26
9.2	Reduction to a quadratic inequality	27
9.3	Spectral form	28
9.4	The sharp L^1 -Khintchine inequality: proof of Theorem K	29
9.5	Proof of Proposition 9.1	33
9.6	A second function: g_1	33
9.7	A level-one inequality for nonnegative functions: proof of Theorem L	35

9.8 A third function: g_2	37
10 A maximal-correlation criterion: proof of Theorem F	38
10.1 An example with two relevant coordinates	39
11 What remains open; numerics	46
A Verification scripts	70

1 Setting and results

1.1 Notation and the CFMMS definitions

We use ± 1 notation. For $g: \{-1, 1\}^m \rightarrow \{-1, 1\}$ we write $g = \sum_{S \subseteq [m]} \hat{g}(S) \prod_{j \in S} x_j$ for its Fourier expansion and identify g with this multilinear polynomial; thus $g(y)$ is defined for $y \in [-1, 1]^m$ (and for $y \in \mathbb{R}^m$). We write

$$\mu = \mu_g = \mathbb{E}g = \hat{g}(\emptyset), \quad \text{triv}_g = \frac{1}{2} + \frac{1}{2}|\mu|, \quad \rho_j = \frac{\hat{g}(\{j\})}{\sqrt{1 - \mu^2}} \quad (j \in [m]).$$

For $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ and Z uniform in $\{-1, 1\}^{n \times m}$, $(f \circ g^n)(Z) = f(g(Z_{1\cdot}), \dots, g(Z_{n\cdot}))$ (rows) and $(g \circ f^m)(Z) = g(f(\text{col}_1 Z), \dots, f(\text{col}_m Z))$ (columns).

Definition 1.1 (CFMMS Def. 5.2). s_g is the infimum of all s with the following property: for every $\epsilon > 0$ there are $\delta > 0$ and $L \in \mathbb{N}$ such that for all n and all $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$, if $\Pr_Z[(f \circ g^n)(Z) = (g \circ f^m)(Z)] \geq s + \epsilon$ then $|\hat{f}(S)| \geq \delta$ for some $|S| \leq L$.

(In CFMMS the displayed event is misprinted as $f \circ g^m = g \circ f^m$.) We call f (δ, L) -pseudorandom if $|\hat{f}(S)| < \delta$ for all $|S| \leq L$. Directly from the definition:

Remark 1.2. (a) If $f_N: \{-1, 1\}^{n_N} \rightarrow \{-1, 1\}$ is a single sequence with $\max_{|S| \leq L} |\hat{f}_N(S)| \rightarrow 0$ for every fixed L and $\liminf_N \Pr[f_N \circ g = g \circ f_N] \geq a$, then $s_g \geq a$. (Indeed, if $s < a$ had the property, take $\epsilon = (a - s)/3$ and the corresponding δ, L ; for large N , f_N is (δ, L) -pseudorandom and has agreement $\geq a - \epsilon > s + \epsilon$.) (b) If for every $\epsilon > 0$ there are δ, L such that every (δ, L) -pseudorandom f has agreement $\leq b + \epsilon$, then $s_g \leq b$ (the value $s = b$ has the property: given ϵ , use (b) with $\epsilon/2$).

Definition 1.3 (CFMMS Def. 5.1). $\mathcal{N}_g$ is the law of a Gaussian vector $(\mathcal{G}_0, \mathcal{G}_1, \dots, \mathcal{G}_m)$ of standard Gaussians such that $\mathcal{G}_1, \dots, \mathcal{G}_m$ are independent and $\mathbb{E}[\mathcal{G}_0 \mathcal{G}_j] = \rho_j$. For $n \in \mathbb{N}$, $\mathcal{N}_g^n$ is the law of $(\mathcal{G}_0, \dots, \mathcal{G}_m) \in (\mathbb{R}^n)^{m+1}$ whose n coordinates are independent copies of $\mathcal{N}_g$.

(This exists since $\sum_j \rho_j^2 = \|g^{\perp}\|_2^2 / (1 - \mu^2) \leq 1$.) For $q_1, \dots, q_m: \mathbb{R}^n \rightarrow [-1, 1]$ measurable put

$$\gamma_q(x) = \mathbb{E}[g(q_1(\mathcal{G}_1), \dots, q_m(\mathcal{G}_m)) \mid \mathcal{G}_0 = x] \quad (x \in \mathbb{R}^n), \quad \text{val}(q) = \frac{1}{2} + \frac{1}{2} \mathbb{E}_{x \sim N(0, I_n)} |\gamma_q(x)|,$$

where the conditional expectation is taken with respect to the explicit Gaussian conditional law of Lemma 2.1 (so γ_q is defined for every x).

Definition 1.4 (CFMMS Thm. 5.3 and Lemma 5.5). $s_g^U = \sup \text{val}(q_1, \dots, q_m)$ over all n and all $q_j: \mathbb{R}^n \rightarrow \{-1, 1\}$ with $\mathbb{E}q_j = 0$; $s_g^L = \sup \text{val}(q, \dots, q)$ over all n and all $q: \mathbb{R}^n \rightarrow [-1, 1]$ with $\mathbb{E}q = 0$.

CFMMS prove, for g depending on all coordinates, not of the form $\pm \prod_j x_j$, and with $\mu \neq 0$: $s_g \leq s_g^U$ (Thm. 5.3), $s_g^U < 1$ (Lemma 5.4), $s_g \geq s_g^L$ (Lemma 5.5), $s_g^L \geq \text{triv}_g$ (discussion after Lemma 5.5; see also Lemma 2.3(a) below); moreover the supremum defining s_g^U is unchanged if $[-1, 1]$ -valued q_j are allowed (CFMMS §5.2.2, “Rounding”, in the proof of Thm. 5.3).

Open questions (CFMMS §7). *Q5: Is $s_g^U = s_g^L$? Is the optimum always achieved in one-dimensional Gaussian space? Q6: For which functions g is $s_g > \max(\mathbb{E}[g], 1 - \mathbb{E}[g])$? (in $\{0, 1\}$ notation; in ± 1 notation the right-hand side is triv_g .)* CFMMS add: “Lemma 5.6 shows that the strict inequality holds for unbalanced g whenever all Fourier coefficients on the first level are non-zero. Conversely, when all Fourier coefficients on the first level vanish, Theorem 5.3 shows that equality holds.”

Overloaded symbols. A few letters are reused with a meaning that is local to one part of the paper. b is a sign in π_b (§7–§9), a $[0, 1]$ -valued test function in Lemma 2.2 and §3, and a coefficient in the proof of Theorem K (§9.4). W is a label matrix (Definition 1.5, §5, §6), whereas $W_k(\cdot)$ (Gaussian Hermite weights, §3) and W^1, W_+ (§9) are Fourier weights. $d = |J|$, except in §6, where (d, τ) are regularity parameters and d_1, d_2 are degree cut-offs. γ_q and γ_ν are conditional expectations of g , whereas in §6 γ, γ_0 are noise rates, and γ_n is the standard Gaussian measure on $\mathbb{R}^n$ (§7, §8). In §4.2, κ, ℓ, σ are functions on $[-1, 1]$ and $c = -\frac{1}{14}$; in the proof of Proposition 4.3, $\kappa(t) = \frac{15+t^2}{14}$, α, β are the constants of that proposition, $L(a)$ is an orthant probability, $X = 12$ is a cut-off and h an integration variable, and in §3 α, β are other constants. In Theorem A, §4 and the paragraphs on E2, $J = \mathbb{E}_x \gamma_q(x)^+$ is a number; elsewhere $J = \{j : \hat{g}(\{j\}) \neq 0\}$. In the proof of Theorem K (§9.4), f is the Stein solution, $Z \sim N(0, 1)$, U is a zero-bias variable, I a random index, $k(\cdot)$ and $m(\cdot)$ are auxiliary functions of the Mills ratio, and $\theta = 1 - 2q$; elsewhere Z is a column of a label matrix, $U = [m] \setminus J$, I is the integral of Theorem A, k the number of labels, m the arity and θ a gap (§7). In §9.7, $\beta = \frac{1-2q}{q}$, $c = \frac{(1-q)^2}{q}$ and h is a nonnegative factor of degree at most 1; in Proposition 10.4 and its proof $\beta = \frac{1-2q}{q}$ again, A is a number, W_1 a weight, $f = \mathbb{E}\bar{P}^2$, and $U(\cdot), V(\cdot)$ are functions of one variable; elsewhere $\beta_\pm$ are biases, $c(j)$ is the class of j and A_w are sets. Φ is the standard Gaussian distribution function, except for the functionals $\Phi, \Phi^+, \Phi^\sharp$ of §10. ρ_j are the Gaussian correlations of the setting, ρ_{ij} maximal correlations (§10), $\rho(L, R)$ a correlation of two variables, and ρ_* the correlation of the coupling in §9.7. In §10, $r = |U|$ in the paragraph on candidates, in Proposition 10.7 and in the paragraphs on six to nine variables; $r = \frac{q}{1-q}$ in Proposition 10.6, its proof and Remark 10.10; and r is a row index in the proofs of Theorem F and of Proposition 10.4. There, too, c is a coefficient in Lemma 10.5 but the less likely value of a ± 1 variable in Remark 10.10, Q_n is a set of vectors (Proposition 10.9), and $\lambda_{\pm 1}$ in Lemma 10.5 is $\lambda_\pm$.

1.2 The labelled value

Definition 1.5 (labelled value). Let $k \geq 0, n \geq 0$ and let $(Q_z)_{z \in \{-1, 1\}^k}$ be measurable functions $Q_z: \mathbb{R}^n \rightarrow [-1, 1]$ with $\mathbb{E}Q_z = 0$ (expectation under $N(0, I_n)$). Let W be uniform in $\{-1, 1\}^{k \times m}$, independent of $(\mathcal{G}_0, \dots, \mathcal{G}_m) \sim \mathcal{N}_g^n$; write $W_{\cdot j} \in \{-1, 1\}^k$ for its columns and $g(W) \in \{-1, 1\}^k$ for g applied to each row. For $\nu \in \{-1, 1\}^k$ put

$$\gamma_\nu(x) = \mathbb{E} \left[g(Q_{W_{\cdot 1}}(\mathcal{G}_1), \dots, Q_{W_{\cdot m}}(\mathcal{G}_m)) \mid \mathcal{G}_0 = x, g(W) = \nu \right],$$

$$\text{val}^{\text{lab}}(Q) = \frac{1}{2} + \frac{1}{2} \sum_\nu \Pr[g(W) = \nu] \mathbb{E}_{x \sim N(0, I_n)} |\gamma_\nu(x)|.$$

s_g^{lab} is the supremum of $\text{val}^{\text{lab}}(Q)$ over all k, n and all such families.

For $k = 0$ there is one function $Q = Q_\emptyset$ and $\text{val}^{\text{lab}}(Q) = \text{val}(Q, \dots, Q)$, so $s_g^L \leq s_g^{\text{lab}}$. (The labels play the role of the influential coordinates of f ; see §5.)

1.3 Results

Theorem A ($g = \text{E2}$). Let $\text{E2}(x_1, x_2, x_3) = +1$ iff exactly two of x_1, x_2, x_3 equal $+1$. Then $\mu = -\frac{1}{4}$, $\hat{g}(\{j\}) = \frac{1}{4}$ for all j , and:

- (a) for every n , every measurable $q: \mathbb{R}^n \rightarrow [-1, 1]$ (mean zero or not) and every $x \in \mathbb{R}^n$, $\gamma_{(q,q,q)}(x) \leq 2 \cdot 0.49788 - 1 < 0$; that is, after randomised rounding of the three inputs, $\Pr[\text{E2} = +1 \mid \mathcal{G}_0 = x] \leq 0.49788 < \frac{1}{2}$. Consequently $s_{\text{E2}}^L = \frac{5}{8} = \text{triv}_{\text{E2}}$;
- (b) $s_{\text{E2}}^U \geq \frac{5}{8} + J > 0.6253166$, where $J = \mathbb{E}_x \gamma_q(x)^+$ for $q = (\text{sgn}, \text{sgn}, -\text{sgn})$ on $\mathbb{R}$ satisfies $J > 3.1663998209 \times 10^{-4}$ (computer-assisted, Proposition 4.3; in floating point $J = 3.16639982095 \times 10^{-4}$); analytically $J \geq I > 0$, where $I = \int_{x_0}^{\infty} (1 - 6\Phi(-x/\sqrt{14}))\varphi(x) dx \approx 1.3588 \times 10^{-5}$ and $x_0 = \sqrt{14}\Phi^{-1}(5/6)$;
- (c) $s_{\text{E2}} \geq s_{\text{E2}}^{\text{lab}} \geq \frac{5}{8} + \frac{3}{8}J > 0.6251187$;
- (d) $s_{\text{E2}} \leq s_{\text{E2}}^U < \frac{5}{8} + 1.95035 \times 10^{-3} < 0.626951$ (computer-assisted; Theorem 4.5).

In particular $s_g^L < s_g^U$ and $s_g^L < s_g$ for $g = \text{E2}$, and CFMMS Lemma 5.6 is false (E2 is unbalanced and all its level-one coefficients are nonzero, but $s_{\text{E2}}^L = \text{triv}_{\text{E2}}$).

Theorem B (exact characterisation). Let g depend on all coordinates, $g \neq \pm \prod_j x_j$, and $\mathbb{E}g \neq 0$. Then $s_g^L \leq s_g = s_g^{\text{lab}} \leq s_g^U$.

For the nontriviality question let $J = \{j \in [m] : \hat{g}(\{j\}) \neq 0\}$, $d = |J|$, $U = [m] \setminus J$, and

$$h(w) = \mathbb{E}_{u \in \{-1, 1\}^U} g(w, u) \quad (w \in \{-1, 1\}^J),$$

so $\hat{h}(S) = \hat{g}(S)$ for $S \subseteq J$, $\mathbb{E}h = \mu$. Call $j, j' \in J$ *equivalent* if $\rho_j = \rho_{j'}$ and let $c(j)$ denote the class of j . For $b \in \{-1, 1\}$ let $\pi_b(w) = \frac{1+bh(w)}{2^d(1+b\mu)}$ (a probability distribution on $\{-1, 1\}^J$: the law of the J -part of a uniform input conditioned on $g = b$). For $k \geq 0$, $\nu \in \{-1, 1\}^k$ and maps $F_c: \{-1, 1\}^k \rightarrow [-1, 1]$ (one per class) put

$$\Gamma_\nu(F) = \mathbb{E} \left[h(F_{c(j)}(C_{\cdot j}))_{j \in J} \right], \quad C \in \{-1, 1\}^{k \times J} \text{ with independent rows, row } i \sim \pi_{\nu_i}.$$

Theorem C (nontriviality). Let g be as in Theorem B. Then:

- (i) if $\text{sgn}(\mu)h(w) \geq 0$ for all $w \in \{-1, 1\}^J$, then $s_g = s_g^U = \text{triv}_g$;
- (ii) if $\text{sgn}(\mu)h(w) < 0$ for some w , then $s_g^U > \text{triv}_g$. Thus $s_g^U > \text{triv}_g$ iff $\exists w : \text{sgn}(\mu)h(w) < 0$;
- (iii) if $\text{sgn}(\mu)\Gamma_\nu(F) < 0$ for some k, ν, F , then $s_g > \text{triv}_g$; if this holds with $k = 0$ (i.e. $\text{sgn}(\mu)h(y) < 0$ for some $y \in [-1, 1]^J$ that is constant on classes), then even $s_g^L > \text{triv}_g$;
- (iv) if $\|h\|_2^2 > |\mu|$, then $s_g > \text{triv}_g$. In particular, if all level-one coefficients of g are nonzero then $s_g > \text{triv}_g$ (the corrected form of CFMMS Lemma 5.6);

- (v) if $m \leq 4$, then $s_g > \text{triv}_g \iff s_g^U > \text{triv}_g \iff \exists w \in \{-1, 1\}^J : \text{sgn}(\mu)h(w) < 0$
(computer-assisted, exact integer arithmetic; §7.1).

Parts (i) and (iv) already decide every g with at most one coordinate outside J , for every m . If $U = \emptyset$, then $h = g$ is ± 1 -valued and nonconstant, so $\|h\|_2^2 = 1 > |\mu|$. If $U = \{u\}$, then $h(w) = \frac{1}{2}(g(w, 1) + g(w, -1)) \in \{-1, 0, 1\}$, so $\|h\|_2^2 = \mathbb{E}|h| \geq |\mathbb{E}h| = |\mu|$, with equality iff $\text{sgn}(\mu)h \geq 0$. Hence, for $|U| \leq 1$, $s_g > \text{triv}_g$ iff $\text{sgn}(\mu)h(w) < 0$ for some w .

The sufficient condition (iii) is in fact also necessary, and the classes play no role:

Theorem D (exact criterion for nontriviality). Let g be as in Theorem B and $\sigma = \text{sgn}(\mu)$. The following are equivalent:

- (a) $s_g > \text{triv}_g$;
- (b) $s_g^{\text{lab}} > \text{triv}_g$;
- (c) $\sigma \Gamma_\nu(F) < 0$ for some $k \geq 0$, $\nu \in \{-1, 1\}^k$ and class-wise maps $F_c: \{-1, 1\}^k \rightarrow [-1, 1]$;
- (d) $\sigma \mathbb{E}[h(F(C_{\cdot j}))_{j \in J}] < 0$ for some $k \geq 0$, $\nu \in \{-1, 1\}^k$ and a *single* map $F: \{-1, 1\}^k \rightarrow [-1, 1]$ (rows of C independent, row $i \sim \pi_{\nu_i}$), which may moreover be taken rational-valued.

In particular, whether s_g is nontrivial depends only on h , and it is certified by a finite exact computation whenever it holds.

The condition $\exists w : \text{sgn}(\mu)h(w) < 0$, which is necessary by Theorem C(i) and equivalent to $s_g^U > \text{triv}_g$ by (ii), is not sufficient once $m \geq 5$:

Theorem E (an explicit g with $s_g = \text{triv}_g < s_g^U$). Let $g_0: \{-1, 1\}^5 \rightarrow \{-1, 1\}$ be the function defined in (10) (§9); it depends on all five coordinates, $\mathbb{E}g_0 = -\frac{1}{2}$, $J = \{1, 2, 3\}$ and $h = (-1, -1, -1, \frac{1}{2}, 0, -\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2})$ in the order $++++, ++-, \dots, ---$. Then

$$s_{g_0}^L = s_{g_0} = \text{triv}_{g_0} = \frac{3}{4} < s_{g_0}^U.$$

Corollary E'. For every $m \geq 5$ there is $g: \{-1, 1\}^m \rightarrow \{-1, 1\}$ depending on all coordinates with $s_g^L = s_g = \text{triv}_g < s_g^U$. For $m \leq 4$ there is no such g with $\mathbb{E}g \neq 0$ (Theorem C(v)).

The restriction $\mathbb{E}g \neq 0$ in the last statement is needed for our argument: Theorems B–D, and hence Theorem C(v), assume $\mu \neq 0$, and for balanced g CFMMS define s_g^U and s_g^L differently; we do not treat that case.

The method of proof of Theorem E also settles a second, inequivalent function: for the function g_1 of (14) (§9.6), with $\mathbb{E}g_1 = -\frac{3}{8}$, one has $s_{g_1}^L = s_{g_1} = \text{triv}_{g_1} = \frac{11}{16} < s_{g_1}^U$ (Proposition 9.10).

The key new ingredient is a sharp L^1 -Khintchine inequality for biased bits, which may be of independent interest. For unbiased bits the corresponding statement is trivial (Cauchy–Schwarz), and for q -biased bits with $q < 1 - \frac{1}{\sqrt{2}} \approx 0.293$ it is false: two equal coefficients, $\alpha = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, give $\mathbb{E}|\chi_1 + \chi_2|/\sqrt{2} = \sqrt{2}(1 - q)\mathbb{E}|\chi_1| > \mathbb{E}|\chi_1|$ (Remark 9.8(c); this example is in Stone’s thesis [16, §4.2]).

Theorem K (sharp L^1 -Khintchine inequality for biased bits). Let $\frac{3}{8} \leq q \leq \frac{5}{8}$ and let $\chi_1, \dots, \chi_n$ be independent with $\Pr[\chi_i = \sqrt{(1-q)/q}] = q$ and $\Pr[\chi_i = -\sqrt{q/(1-q)}] = 1 - q$ (so $\mathbb{E}\chi_i = 0$, $\mathbb{E}\chi_i^2 = 1$). Then for every $\alpha \in \mathbb{R}^n$ with $\sum_i \alpha_i^2 = 1$,

$$\mathbb{E}\left|\sum_{i=1}^n \alpha_i \chi_i\right| \leq \mathbb{E}|\chi_1| = 2\sqrt{q(1-q)}.$$

Equivalently, every $f: \{-1, 1\}^n \rightarrow [0, 1]$ has level-one weight $W^1(f) \leq q(1 - q)$ with respect to the product measure with $\Pr[x_i = 1] = q$, with equality for $f(x) = \mathbf{1}\{x_1 = 1\}$.

The proof (§9.4) is by induction on n : a zero-bias Stein bound handles vectors whose coefficients are all small (at most $\frac{3}{10}$ for $q = \frac{3}{8}$), and explicit quadratic majorants handle the inductive step otherwise. The argument is uniform in q except for two inequalities in one or two real variables, which we verify by hand for $q = \frac{3}{8}, \frac{2}{5}, \frac{5}{12}$ and by an exact rational interval computation for all $q \in [\frac{3}{8}, \frac{1}{2}]$; the case $q > \frac{1}{2}$ follows by the symmetry $\chi_i \mapsto -\chi_i$. Theorem E uses $q = \frac{3}{8}$ and Proposition 9.10 below uses $q = \frac{2}{5}$. The range cannot be extended to all q with $q(1 - q) \geq \frac{1}{2\pi}$ (the range in which the central limit theorem gives no obstruction): for $q < 1 - \frac{1}{\sqrt{2}}$ the two-coefficient example above violates the inequality (Remark 9.8). In the notation of Ivanisvili and Stone [10], who study the best constant $R(q)$ in $\mathbb{E}|L| \leq R(q)(\mathbb{E}L^2)^{1/2}$ for such sums in connection with the L^1 -Poincaré inequality on the Hamming cube [9], Theorem K determines $R(q) = 2\sqrt{q(1 - q)}$ for $q \in \{\frac{3}{8}, \frac{2}{5}, \frac{5}{12}\}$ and the mirror values $\frac{5}{8}, \frac{3}{5}, \frac{7}{12}$, and (computer-assisted) for all $q \in [\frac{3}{8}, \frac{5}{8}]$; see Remark 9.8(d) for the literature.

A third boundary function needs a level-one bound that depends on the mean, and for it a short spectral argument suffices (§9.7). It is of Hoffman type, as in Friedgut's proof of the biased Erdős–Ko–Rado theorem [5].

Theorem L (level-one inequality for nonnegative functions). Let $0 < q \leq \frac{1}{2}$ and let $f: \{-1, 1\}^n \rightarrow [0, \infty)$, where $\{-1, 1\}^n$ carries the product measure with $\Pr[x_i = 1] = q$. Then

$$W^1(f) \leq q \mathbb{E}[f^2] + \frac{1 - 2q}{q} (\mathbb{E}f)^2.$$

In particular $W^1(f) \leq qp + \frac{1 - 2q}{q} p^2$ if f takes values in $[0, 1]$ and $\mathbb{E}f = p$. For $q < \frac{1}{2}$, equality holds in the first inequality iff $\hat{f}(S) = 0$ for all $|S| \geq 3$ and $f(x)f(y) = 0$ for all x, y whose sets of $+1$ coordinates are disjoint, that is, iff up to a permutation of the coordinates

$$f = \mathbf{1}\{x_1 = 1\} h(x_2, \dots, x_n) \quad \text{or} \quad f = a \mathbf{1}\{x_1 = x_2 = 1\} + b \mathbf{1}\{x_1 = x_3 = 1\} + c \mathbf{1}\{x_2 = x_3 = 1\},$$

where $h \geq 0$ has degree at most 1 and $a, b, c \geq 0$. For $q = \frac{1}{2}$, equality holds iff $\hat{f}(S) = 0$ for all odd $|S| \geq 3$ and $f(x)f(-x) = 0$ for all x . Equality in the second inequality holds iff in addition f is $\{0, 1\}$ -valued; for $q < \frac{1}{2}$ this leaves $f = 0$ and, up to a permutation of the coordinates, $\mathbf{1}\{x_1 = 1\}$, $\mathbf{1}\{x_1 = x_2 = 1\}$ and $\mathbf{1}\{x_1 = 1, x_2 = -1\}$.

With Theorem L the method also settles the function g_2 of (17) (§9.8): $\mathbb{E}g_2 = -\frac{3}{8}$ and $s_{g_2}^L = s_{g_2} = \text{triv}_{g_2} = \frac{11}{16} < s_{g_2}^U$ (Proposition 9.11); here the dictator decoder and the AND of two labels are both extremal.

When h has no terms of degree ≥ 3 there is a soft criterion for triviality, based on the maximal correlation of the columns of the label matrix (§10). For distinct $i, j \in J$ and $b \in \{-1, 1\}$ let $\text{Corr}_{\pi_b}(w_i, w_j)$ be the correlation of the coordinates w_i, w_j under π_b (defined as 0 if one of them is π_b -a.s. constant), and put $\rho_{ij} = \max_b |\text{Corr}_{\pi_b}(w_i, w_j)|$.

Theorem F (maximal-correlation criterion). Let g be as in Theorem B, $\sigma = \text{sgn}(\mu)$, and suppose that $\hat{g}(S) = 0$ for every $S \subseteq J$ with $|S| \geq 3$ (automatic if $|J| \leq 2$). If

$$\Phi(y) := \sigma h(y) - \sum_{\{i, j\} \subseteq J} |\hat{h}(\{i, j\})| \rho_{ij} \sqrt{(1 - y_i^2)(1 - y_j^2)} \geq 0$$

for every $y \in [-1, 1]^J$ that is constant on classes, then $\sigma \Gamma_\nu(F) \geq 0$ for all k, ν and class-wise F , and hence $s_g^L = s_g = \text{triv}_g$.

Theorem F can only help inside classes: at a vertex w that is constant on classes the square roots vanish, so the hypothesis contains $\sigma h(w) \geq 0$, and if $\sigma h(w) < 0$ at such a vertex then Theorem C(iii) with $k = 0$ already gives $s_g^L > \text{triv}_g$ (Corollary 10.2). With two relevant coordinates in one class it gives a new kind of example (§10.1): the function g_F on 6 variables defined in (18) has $J = \{1, 2\}$, $\hat{g}(\{1\}) = \hat{g}(\{2\})$, $\mathbb{E}g_F = -\frac{1}{8}$, and

$$s_{g_F}^L = s_{g_F} = \text{triv}_{g_F} = \frac{9}{16} < s_{g_F}^U.$$

Corollary F'. For every $m \geq 6$ there is $g: \{-1, 1\}^m \rightarrow \{-1, 1\}$ depending on all coordinates with $|J| = 2$ and $s_g^L = s_g = \text{triv}_g < s_g^U$. For $m \leq 5$ there is no such g with $\mathbb{E}g \neq 0$ and $|J| \leq 2$ (computer-assisted: Theorem C(v) and Proposition 11.1).

For one class of size two, Proposition 10.4 sharpens Theorem F by applying Theorem L to the part of the covariance that comes from single rows of the more dangerous sign, and Proposition 10.6 sharpens it further by comparing the column pair with a second coupling (the anti-coupling of Theorem L in the dangerous rows). Together with an exact enumeration of the possible h (Proposition 10.7) this settles all cases with $|J| \leq 2$ and $m \leq 8$: $s_g > \text{triv}_g$ iff Theorem D(c) holds with $k \leq 3$, and $k \leq 2$ suffices for $m \leq 7$ (Corollary 10.8). For $m = 8$ it does not: the function g_3 of Proposition 10.9 has $s_{g_3} > \text{triv}_{g_3}$, but its smallest certificate has $k = 3$.

Answers. Q5, first part: no, $s_g^U \neq s_g^L$ in general (Theorem A), and the true value is the intermediate labelled value (Theorem B); the CFMMS upper bound is not tight either, $s_g < s_g^U$ can happen (Theorem E). The second part of Q5 (whether the optimum is attained in one-dimensional Gaussian space) remains open (§11). Q6 (for unbalanced g): Theorem D characterises nontriviality exactly by finite “labelled decoder” certificates, a condition on h alone that can be certified by a finite exact computation whenever it holds (no bound on k is known; §11). Theorem C gives a necessary condition, which is also sufficient for $m \leq 4$ (C(v)), and sufficient conditions (C(iii), (iv)); by Theorem E and Corollary E' the necessary condition is not sufficient for any $m \geq 5$. It is sufficient whenever the nonzero level-one coefficients of g are pairwise distinct (Corollary 10.2): the failures come from coordinates in the same class. With only two relevant coordinates they first occur at $m = 6$ (Theorem F, Corollary F'), and for $|J| \leq 2$ and $m \leq 8$ nontriviality is decided exactly, by certificates with $k \leq 3$ on one side and Theorem F and Propositions 10.4 and 10.6 on the other (Corollary 10.8). What remains open is described in §11.

Numerical values for E2. By Theorem A, $0.6251187 < s_{E2} \leq s_{E2}^U < 0.626951$ and $s_{E2}^U > 0.6253166$. The lower bounds are the values $\text{val}(\text{sgn}, \text{sgn}, -\text{sgn}) = \frac{5}{8} + J$ and $\frac{5}{8} + \frac{3}{8}J \leq \text{val}^{\text{lab}}(z \text{sgn})$ (one label, $n = 1$); floating-point evaluation gives $\text{val}^{\text{lab}}(z \text{sgn}) - \frac{5}{8} - \frac{3}{8}J \approx 5.7 \cdot 10^{-9}$ (`e2_check.py`, and the same value with the quadrature of `e2_label_roles.py`), so the second output label contributes almost nothing. By Theorem A(d), $s_{E2}^U - \frac{5}{8} < 1.9504 \times 10^{-3}$, within a factor 6.2 of $J = 3.1664 \times 10^{-4}$; the bound $s_{E2}^U \leq \frac{1}{2} + \frac{1}{2}\sqrt{\langle T_{\rho'}g, g \rangle} = 0.650666\dots$ contained in the proof of CFMMS Lemma 5.4 ($\rho' = \frac{2}{\pi} \arcsin \frac{1}{5}$, $\langle T_{\rho'}g, g \rangle = \frac{1}{16}(1 + 3\rho' + 3\rho'^2 + 9\rho'^3)$) has excess 2.567×10^{-2} . The exact value of s_{E2} remains unknown.

2 Preliminaries

Multilinear facts. The following apply to g and to any multilinear polynomial bounded by 1 on $\{-1, 1\}^m$, e.g. to h .

(M1) If $Y_1, \dots, Y_m$ are independent bounded random variables, then expanding the multilinear polynomial gives $\mathbb{E}g(Y_1, \dots, Y_m) = g(\mathbb{E}Y_1, \dots, \mathbb{E}Y_m)$. In particular, for $y \in [-1, 1]^m$ we

have $g(y) = \mathbb{E}g(Y)$, where the $Y_j \in \{-1, 1\}$ are independent with $\mathbb{E}Y_j = y_j$ (“randomised rounding”); so $|g(y)| \leq 1$.

(M2) For $y, y' \in [-1, 1]^m$, $|g(y) - g(y')| \leq \sum_j |y_j - y'_j|$: g is affine in each variable and $\partial_j g(y) = \frac{1}{2}(g(y^{j \rightarrow 1}) - g(y^{j \rightarrow -1})) \in [-1, 1]$; change one coordinate at a time.

Lemma 2.1 (conditional law). *Let $(\mathcal{G}_0, \dots, \mathcal{G}_m) \sim \mathcal{N}_g^n$. Conditionally on $\mathcal{G}_0 = x \in \mathbb{R}^n$, $(\mathcal{G}_1, \dots, \mathcal{G}_m)$ is Gaussian with mean $(\rho_1 x, \dots, \rho_m x)$ and covariance $(I_m - \rho \rho^\top) \otimes I_n$. If $\rho_j = 0$ then $\mathcal{G}_j$ is independent of $(\mathcal{G}_0, (\mathcal{G}_i)_{i \neq j})$.*

Proof. Standard Gaussian conditioning: $\mathbb{E}[\mathcal{G}_j \mid \mathcal{G}_0] = \rho_j \mathcal{G}_0$ and $\text{Cov}(\mathcal{G}_i, \mathcal{G}_j \mid \mathcal{G}_0) = \delta_{ij} - \rho_i \rho_j$, coordinatewise. If $\rho_j = 0$, $\mathcal{G}_j$ is uncorrelated with, hence independent of, the jointly Gaussian vector $(\mathcal{G}_0, (\mathcal{G}_i)_{i \neq j})$. $\square$

Gaussian analysis. Let $\gamma_n = N(0, I_n)$. Every $\phi \in L^2(\gamma_n)$ has an orthogonal (Hermite / Wiener chaos) decomposition $\phi = \sum_{k \geq 0} \phi^{=k}$; write $W_k(\phi) = \|\phi^{=k}\|_2^2$. Then $\phi^{=0} = \mathbb{E}\phi$, $\phi^{=1}(y) = \langle \mathbb{E}[\phi(Y)Y], y \rangle$, so $W_1(\phi) = \|\mathbb{E}[\phi(Y)Y]\|_2^2$, and $\sum_{k \geq 1} W_k(\phi) = \text{Var } \phi$. If Y, Y' are jointly Gaussian standard Gaussian vectors in $\mathbb{R}^n$ with $\mathbb{E}[Y_r Y'_s] = r_0 \delta_{rs}$ ($|r_0| \leq 1$), then (Mehler)

$$\mathbb{E}[\phi(Y)\psi(Y')] = \sum_{k \geq 0} r_0^k \langle \phi^{=k}, \psi^{=k} \rangle, \quad \mathbb{E}[\psi(Y') \mid Y] = (U_{r_0} \psi)(Y), \quad U_{r_0} \psi = \sum_k r_0^k \psi^{=k} \quad (1)$$

(see e.g. O’Donnell, *Analysis of Boolean Functions*, Ch. 11).

Lemma 2.2 (level-one bound). *If $b: \mathbb{R}^n \rightarrow [0, 1]$ is measurable then $W_1(b) \leq \varphi(\Phi^{-1}(\mathbb{E}b))^2 \leq \frac{1}{2\pi}$.*

Proof. If $\mathbb{E}[bY] = 0$ there is nothing to prove. Otherwise let $e = \mathbb{E}[bY] / \|\mathbb{E}[bY]\|$ and $X = \langle e, Y \rangle \sim N(0, 1)$. Then $\|\mathbb{E}[bY]\| = \mathbb{E}[b(Y)X]$. Among measurable $0 \leq b' \leq 1$ with $\mathbb{E}b' = p := \mathbb{E}b$, $\mathbb{E}[b'X]$ is maximised by $b' = \mathbf{1}\{X > t\}$ with $\Pr[X > t] = p$ (bathtub principle: $\mathbb{E}[(b - b')(X - t)] \leq 0$), giving $\int_t^\infty s \varphi(s) ds = \varphi(t) = \varphi(\Phi^{-1}(p)) \leq (2\pi)^{-1/2}$. $\square$

Elementary facts about the labelled value.

Lemma 2.3. *Let (Q_z) be as in Definition 1.5.*

(a) $\mathbb{E}_x \gamma_\nu(x) = \mu$ for every ν . Hence $\mathbb{E}_x |\gamma_\nu| = |\mu| + 2\mathbb{E}_x[(\text{sgn}(\mu)\gamma_\nu)^-]$ and

$$\text{val}^{\text{lab}}(Q) = \text{triv}_g + \sum_\nu \Pr[g(W) = \nu] \mathbb{E}_x[(\text{sgn}(\mu)\gamma_\nu)^-],$$

where $t^- = \max(-t, 0)$. So $\text{val}^{\text{lab}}(Q) > \text{triv}_g$ iff some γ_ν has sign opposite to μ on a set of positive measure. The same holds for $\text{val}(q)$ with γ_q (for mean-zero q_j). In particular $s_g^L, s_g^{\text{lab}}, s_g^U \geq \text{triv}_g$.

(b) (Reduction to h.) $\gamma_\nu(x) = \mathbb{E}[h((Q_{C,j}(\mathcal{G}_j))_{j \in J}) \mid \mathcal{G}_0 = x]$, where $C \in \{-1, 1\}^{k \times J}$ is independent of the $\mathcal{G}$ ’s with independent rows, row $i \sim \pi_{\nu_i}$; indeed, conditionally on $g(W) = \nu$, the J -columns $W_{\cdot j}$ have the law of C . Similarly $\gamma_q(x) = \mathbb{E}[h((q_j(\mathcal{G}_j))_{j \in J}) \mid \mathcal{G}_0 = x]$ for mean-zero q_j .

(c) $s_g^L \leq s_g^{\text{lab}} \leq s_g^U$.

Proof. (a) Given W , the vectors $\mathcal{G}_1, \dots, \mathcal{G}_m$ are independent standard Gaussians (unconditionally on $\mathcal{G}_0$), so the inputs $Q_{W_j}(\mathcal{G}_j)$ are independent with mean 0 and by (M1) $\mathbb{E}[g(\dots) \mid W] = g(0, \dots, 0) = \mu$. Averaging over W with $g(W) = \nu$ gives $\mathbb{E}_x \gamma_\nu = \mu$. Then $|t| = \text{sgn}(\mu)t + 2(\text{sgn}(\mu)t)^-$ gives the formula. (b) Fix W and x . By Lemma 2.1, for $j \in U$ the $\mathcal{G}_j$ are independent of each other and of $(\mathcal{G}_0, (\mathcal{G}_i)_{i \in J})$, so given W , x and $(\mathcal{G}_i)_{i \in J}$ the inputs $Q_{W_j}(\mathcal{G}_j)$, $j \in U$, are independent with mean $\mathbb{E}Q_{W_j} = 0$; by (M1) we may replace them by 0, and $g(y_J, 0_U) = h(y_J)$. Finally the J -columns of W conditioned on $g(W) = \nu$ have independent rows (the rows of W are independent and $g(W)_i$ depends on row i only), and row i has law $\Pr[w \mid g(w, u) = \nu_i] = \frac{2^{-d(1+\nu_i h(w))/2}}{(1+\nu_i \mu)^{d/2}} = \pi_{\nu_i}(w)$. (c) The first inequality is the case $k = 0$. For the second, for fixed W let $q_j^{(W)} = Q_{W_j}$; then $\gamma_\nu = \mathbb{E}[\gamma_{q^{(W)}} \mid g(W) = \nu]$, so by the triangle inequality $\sum_\nu \Pr[\nu] \mathbb{E}|\gamma_\nu| \leq \mathbb{E}_W \mathbb{E}_x |\gamma_{q^{(W)}}(x)| \leq 2s_g^U - 1$, using that s_g^U is unchanged when $[-1, 1]$ -valued functions are allowed (CFMMS §5.2.2). $\square$

3 Theorem A(a): $s_{\mathbb{E}2}^L = 5/8$

For E2: $\mathbb{E}2(x) = -\frac{1}{4} + \frac{1}{4} \sum_j x_j - \frac{1}{4} \sum_{i < j} x_i x_j - \frac{3}{4} x_1 x_2 x_3$, so $\mu = -\frac{1}{4}$, $\hat{g}(\{j\}) = \frac{1}{4}$, $\rho_j = \rho = \frac{1/4}{\sqrt{15/4}} = \frac{1}{\sqrt{15}}$ for $j = 1, 2, 3$, and $\text{triv} = \frac{5}{8}$.

Lemma 3.1. *Let $c \in (-\frac{1}{2}, 0]$, $c' = \frac{c}{1+c}$, and let Y_1, Y_2, Y_3 be jointly Gaussian standard Gaussian vectors in $\mathbb{R}^n$ such that the n coordinates of (Y_1, Y_2, Y_3) are i.i.d. and $\mathbb{E}[Y_{i,r} Y_{j,r}] = c$ for $i \neq j$. For every measurable $b: \mathbb{R}^n \rightarrow [0, 1]$ with $p = \mathbb{E}b(Y_1)$,*

$$3 \mathbb{E}[b(Y_1)b(Y_2)(1 - b(Y_3))] \leq 3p^2(1 - p) + 6|c|W_1(b)p + (6|c|^3p + 3c^2 + 3c'^2)p(1 - p).$$

Proof. Conditionally on $Y_3 = y$ we have $Y_i = cy + \sqrt{1 - c^2}Z_i$ ($i = 1, 2$), where (Z_1, Z_2) is independent of Y_3 , $Z_i \sim N(0, I_n)$ and $\mathbb{E}[Z_{1,r} Z_{2,s}] = c' \delta_{rs}$ (indeed $\text{Cov}(Y_1, Y_2 \mid Y_3) = c - c^2$ and $\text{Var}(Y_i \mid Y_3) = 1 - c^2$). Put $b_y(z) = b(cy + \sqrt{1 - c^2}z)$ and $v(y) = \mathbb{E}b_y(Z_1) = (U_c b)(y)$. By (1) with $r_0 = c' \leq 0$, using that the odd- k terms are ≤ 0 and $c'^k \leq c'^2$ for even $k \geq 2$, and $b_y^2 \leq b_y$:

$$\mathbb{E}[b(Y_1)b(Y_2) \mid Y_3 = y] = \sum_k c'^k W_k(b_y) \leq v(y)^2 + c'^2(\mathbb{E}b_y^2 - v(y)^2) \leq v(y)^2 + c'^2 v(y)(1 - v(y)).$$

Multiplying by $1 - b(y) \in [0, 1]$ and integrating over $y \sim N(0, I_n)$,

$$\begin{aligned} \mathbb{E}[b(Y_1)b(Y_2)(1 - b(Y_3))] &\leq \mathbb{E}[(1 - b)v^2] + c'^2 \mathbb{E}[v(1 - v)] \\ &\leq \mathbb{E}[(1 - b)v^2] + c'^2 p(1 - p), \end{aligned}$$

since $\mathbb{E}v = p$ and $\mathbb{E}v^2 \geq p^2$. Write $v = p + w$ with $w = \sum_{k \geq 1} c^k b^k$, $\mathbb{E}w = 0$. Then

$$\mathbb{E}[(1 - b)v^2] = p^2(1 - p) + 2p \mathbb{E}[(1 - b)w] + \mathbb{E}[(1 - b)w^2].$$

Here, as $c \leq 0$ and $\sum_{k \geq 1} W_k(b) = \text{Var } b \leq p(1 - p)$,

$$\mathbb{E}[(1 - b)w] = -\mathbb{E}[bw] = \sum_{k \geq 1} (-1)^{k+1} |c|^k W_k(b) \leq |c|W_1(b) + |c|^3 \sum_{k \geq 3} W_k(b) \leq |c|W_1(b) + |c|^3 p(1 - p),$$

and $\mathbb{E}[(1 - b)w^2] \leq \mathbb{E}w^2 = \sum_{k \geq 1} c^{2k} W_k(b) \leq c^2 p(1 - p)$. Collecting terms and multiplying by 3 proves the claim. $\square$

Theorem 3.2. Let $g = \text{E2}$, $n \geq 1$, $q: \mathbb{R}^n \rightarrow [-1, 1]$ measurable, and $x \in \mathbb{R}^n$. Then $\frac{1+\gamma_{(q,q,q)}(x)}{2} \leq \max_{p \in [0,1]} \tilde{B}(p) < 0.49788$, where

$$\tilde{B}(p) = 3p^2(1-p) + \frac{3}{14\pi}p + \left(\frac{3}{1372}p + \frac{3}{196} + \frac{3}{169}\right)p(1-p).$$

Consequently $\gamma_{(q,q,q)} < 0$ everywhere, $\text{val}(q, q, q) = \frac{5}{8}$ for every mean-zero q , and $s_{\text{E2}}^L = \frac{5}{8}$.

Proof. Let $a = (1+q)/2 \in [0, 1]$. The multilinear polynomial $(1+\text{E2}(2t-1))/2$ in $t \in \{0, 1\}^3$ equals $\Theta(t) = t_1t_2(1-t_3) + t_1t_3(1-t_2) + t_2t_3(1-t_1)$. By (M1) (randomised rounding), for $y \in [-1, 1]^3$, $\frac{1+\text{E2}(y)}{2} = \Theta(\frac{1+y}{2})$. By Lemma 2.1, given $\mathcal{G}_0 = x$ we can write $\mathcal{G}_j = \rho x + \sqrt{1-\rho^2}Y_j$ with (Y_1, Y_2, Y_3) as in Lemma 3.1 with $c = \frac{-\rho^2}{1-\rho^2} = -\frac{1}{14}$ (so $c' = -\frac{1}{13}$). With $b(y) = a(\rho x + \sqrt{1-\rho^2}y)$ and the symmetry of the law of (Y_1, Y_2, Y_3) ,

$$\frac{1+\gamma_{(q,q,q)}(x)}{2} = \mathbb{E}[\Theta(b(Y_1), b(Y_2), b(Y_3))] = 3\mathbb{E}[b(Y_1)b(Y_2)(1-b(Y_3))].$$

Lemma 3.1 with Lemma 2.2 ($W_1(b) \leq \frac{1}{2\pi}$), $6|c|\frac{1}{2\pi} = \frac{3}{14\pi}$, $6|c|^3 = \frac{3}{1372}$, $3c^2 = \frac{3}{196}$, $3c'^2 = \frac{3}{169}$, gives the bound $\tilde{B}(p)$ with $p = \mathbb{E}b(Y_1) \in [0, 1]$.

It remains to show $\max_{[0,1]} \tilde{B} < 0.49788$. The coefficient of $1/\pi$ in $\tilde{B}$ is $\frac{3}{14}p \geq 0$ and $\pi > 3.14$, so $\tilde{B} \leq \bar{B}$ on $[0, 1]$, where $\bar{B}$ is $\tilde{B}$ with $1/\pi$ replaced by $\frac{50}{157}$:

$$\frac{1}{2} - \bar{B}(p) = \frac{4119}{1372}p^3 - \frac{344223}{115934}p^2 - \frac{526815}{5200468}p + \frac{1}{2}.$$

This cubic has no real root in $[0, 1]$ (exact Sturm-sequence count) and equals $\frac{1}{2}$ at $p = 0$, so it is positive on $[0, 1]$; this is all that is needed for $s_{\text{E2}}^L = \frac{5}{8}$. For the constant 0.49788: the cubic $\frac{49788}{100000} - \bar{B}(p)$ likewise has no real root in $[0, 1]$ (exact Sturm count) and equals $\frac{12447}{25000} > 0$ at $p = 0$, so $\bar{B} < 0.49788$ on $[0, 1]$. The maximum of $\bar{B}$ on $[0, 1]$ is attained at the critical point $p_* = \frac{76494}{232037} + \frac{\sqrt{159156415444069}}{36429809} \approx 0.675965$, where $\bar{B}(p_*) \approx 0.4978777$ (both counts and this value: `script numerics/e2_rigorous.py`). For a hand check: with $\alpha = \frac{3}{14\pi} < 0.06823$ and $\beta = \frac{3}{1372} + \frac{3}{196} + \frac{3}{169} < 0.035245$ we have $\tilde{B}(p) \leq \phi(p) := 3p^2(1-p) + \alpha p + \beta p(1-p)$; $\phi'(p) = 0$ iff $9p^2 - (6-2\beta)p - (\alpha + \beta) = 0$, whose root in $[0, 1]$ is $p \approx 0.67585$, where $\phi \approx 0.49802 < \frac{1}{2}$, and $\phi(0) = 0$, $\phi(1) = \alpha$.

Hence $\gamma_{(q,q,q)}(x) \leq 2 \cdot 0.49788 - 1 < 0$ for all x ; by Lemma 2.3(a) (case $k = 0$), $\text{val}(q, q, q) = \text{triv} = \frac{5}{8}$ for every mean-zero q . $\square$

Remark 3.3 (where the CFMMS proof of Lemma 5.6 fails). CFMMS prove Lemma 5.6 with $n = 1$, $q = \text{sgn}$: they note $\mathbb{E}|\gamma| \geq |\mathbb{E}\gamma| = |\mathbb{E}g|$ and assert that equality holds “only if $|\gamma(x)| = |\mathbb{E}g|$ almost surely”, then observe $|\gamma(x)| \rightarrow 1$ as $|x| \rightarrow \infty$. But equality in $\mathbb{E}|\gamma| \geq |\mathbb{E}\gamma|$ holds if and only if γ does not change sign. For E2 all $\rho_j > 0$, so $\gamma(x) \rightarrow \text{E2}(+, +, +) = -1$ as $x \rightarrow +\infty$ and $\gamma(x) \rightarrow \text{E2}(-, -, -) = -1$ as $x \rightarrow -\infty$; by Theorem 3.2, $\gamma < 0$ everywhere, so there is no gain although $|\gamma| \rightarrow 1$. In general, the gain requires $\text{sgn}(\mu)\gamma < 0$ somewhere (Lemma 2.3(a)), and $q = \text{sgn}$ only reaches the two points $\pm(\text{sgn } \rho_j)_j$ of the cube in the limit.

Remark 3.4. The bound is not an artefact of $n = 1$ or of Boolean q : it holds for all n , all $[-1, 1]$ -valued q , and even without the constraint $\mathbb{E}q = 0$. The sharper bound obtained by keeping $W_1(b) \leq \varphi(\Phi^{-1}(p))^2$ has maximum 0.4900295 (at $p \approx 0.66016$). For $q = \text{sgn}$ ($n = 1$) the true maximum of $\Pr[\text{E2} = +1 \mid \mathcal{G}_0 = x]$ is 0.47363 (numerical).

4 Theorem A(b)–(d): bounds for s_{E2}^U and s_{E2}

4.1 Lower bounds

Proposition 4.1. *For $g = E2$, $n = 1$ and $(q_1, q_2, q_3) = (\text{sgn}, \text{sgn}, -\text{sgn})$ we have $\gamma_q(x) \geq 1 - 6\Phi(-x/\sqrt{14})$ for all $x \in \mathbb{R}$. Hence $s_{E2}^U \geq \text{val}(q) \geq \frac{5}{8} + I$ with I as in Theorem A, and $I > 0$.*

Proof. Given $\mathcal{G}_0 = x$, $\mathcal{G}_j \sim N(\rho x, 1 - \rho^2)$, so $\Pr[\mathcal{G}_j \leq 0 \mid x] = \Phi(-\rho x/\sqrt{1 - \rho^2}) = \Phi(-x/\sqrt{14})$. On the event $\{\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3 > 0\}$ the inputs are $(+1, +1, -1)$ and $E2 = +1$; by the union bound $\Pr[E2 = +1 \mid x] \geq 1 - 3\Phi(-x/\sqrt{14})$, i.e. $\gamma_q(x) \geq 1 - 6\Phi(-x/\sqrt{14})$, which is positive for $x > x_0$. The q_j are ± 1 -valued with mean 0, and by Lemma 2.3(a), $\text{val}(q) = \frac{5}{8} + \mathbb{E}\gamma_q^+ \geq \frac{5}{8} + \int_{x_0}^{\infty} (1 - 6\Phi(-x/\sqrt{14}))\varphi(x) dx$. The integrand is positive on (x_0, ∞) , so $I > 0$; numerically $I = 1.3588002 \times 10^{-5}$ (mpmath, 30 digits). $\square$

Proposition 4.2. *For $g = E2$, $k = 1$, $n = 1$ and $Q_z(y) = z \text{sgn}(y)$ ($z \in \{-1, 1\}$) we have $\text{val}^{\text{lab}}(Q) \geq \frac{5}{8} + \frac{3}{8}\mathbb{E}\gamma_q^+ \geq \frac{5}{8} + \frac{3}{8}I$, where $q = (\text{sgn}, \text{sgn}, -\text{sgn})$. Hence $s_{E2} \geq s_{E2}^{\text{lab}} > \frac{5}{8}$.*

Proof. $\mathbb{E}Q_z = 0$. Here $W = (w_1, w_2, w_3)$ is a single uniform row and $\Pr[g(W) = +1] = \frac{3}{8}$. Conditioned on $g(W) = +1$, W is uniform on $\{(+, +, -), (+, -, +), (-, +, +)\}$ and the inputs are $(w_j \text{sgn}(\mathcal{G}_j))_j$. Given $\mathcal{G}_0 = x$ the variables $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$ are exchangeable and $E2$ is symmetric, so each of the three patterns contributes $\gamma_q(x)$, i.e. $\gamma_+ = \gamma_q$. By Lemma 2.3(a), $\text{val}^{\text{lab}}(Q) \geq \frac{5}{8} + \frac{3}{8}\mathbb{E}\gamma_+^+ = \frac{5}{8} + \frac{3}{8}\mathbb{E}\gamma_q^+$, and $\mathbb{E}\gamma_q^+ \geq I$ by Proposition 4.1. The last statement follows from Theorem 5.1 below. $\square$

Proposition 4.3 (computer-assisted). *Let $q = (\text{sgn}, \text{sgn}, -\text{sgn})$ on $\mathbb{R}$, $J = \mathbb{E}_x \gamma_q(x)^+$, $\beta = \sqrt{15/13}$, $\alpha = \sqrt{7/6}$, and let $T(h, s) = \frac{1}{2\pi} \int_0^s e^{-h^2(1+t^2)/2} \frac{dt}{1+t^2}$ be Owen's function. Then, with $a = x/\sqrt{14}$,*

$$\gamma_q(x) = 6L(a) - 4\Phi(a) + 16T(a, \beta) - 1, \quad L(a) = 3 \int_{-\infty}^a \varphi(h)(\Phi(\beta h) - 2T(\beta h, \alpha)) dh, \quad (2)$$

and $J > 3.1663998209 \times 10^{-4}$. Consequently $s_{E2}^U \geq \frac{5}{8} + J > 0.6253166$ and $s_{E2} \geq \frac{5}{8} + \frac{3}{8}J > 0.6251187$.

Proof. Given $\mathcal{G}_0 = x$ write $\mathcal{G}_j = \rho x + \sqrt{1 - \rho^2} Y_j$, where (Y_1, Y_2, Y_3) is standard Gaussian with pairwise correlations $c = -\frac{1}{14}$ (Lemma 2.1); then $\mathcal{G}_j > 0$ iff $Y_j > -a$. Let $A_j = \{Y_j > -a\}$. The inputs $(\text{sgn } \mathcal{G}_1, \text{sgn } \mathcal{G}_2, -\text{sgn } \mathcal{G}_3)$ have exactly two entries $+1$ on $A_1 A_2 A_3$, on $A_1 A_2^c A_3^c$ and on $A_1^c A_2 A_3^c$, so by exchangeability and inclusion–exclusion $\Pr[E2 = 1 \mid x] = L + 2(\Phi(a) - 2\Phi_2(a, a; c) + L)$, where $L = \Pr[A_1 A_2 A_3]$ and $\Phi_2(\cdot, \cdot; r)$ is the bivariate normal distribution function with correlation r (as $(-Y_j)_j$ has the law of $(Y_j)_j$, $\Pr[A_1 A_2] = \Phi_2(a, a; c)$). Owen's formula $\Phi_2(h, h; r) = \Phi(h) - 2T(h, \sqrt{(1-r)/(1+r)})$ and $\sqrt{(1-c)/(1+c)} = \beta$ give (2) for $\gamma_q = 2\Pr[E2 = 1 \mid x] - 1$, once L is identified. Now $L = \Pr[Y_j < a \ \forall j]$ tends to 0 as $a \rightarrow -\infty$, and $\frac{d}{da} \Pr[Y_j < a \ \forall j] = 3\varphi(a) \Pr[Y_1, Y_2 < a \mid Y_3 = a]$. Given $Y_3 = a$ the pair (Y_1, Y_2) has means ca , variances $1 - c^2$ and correlation $c' = c/(1+c) = -\frac{1}{13}$, so the conditional probability is $\Phi_2(\beta a, \beta a; c')$, and Owen's formula with $\sqrt{(1-c')/(1+c')} = \alpha$ gives the formula for L .

Since $\gamma_q^+ \geq \gamma_q$, $J \geq \int_{x_1}^X \gamma_q \varphi dx$ for all $x_1 < X$. We take $X = 12$ and $x_1 = 128826889/(5 \cdot 10^7)$, close to the zero 2.5765378 of γ_q (floating point; any x_1 gives a valid bound). Put $h_1 = x_1/\sqrt{14}$, $H = X/\sqrt{14}$, $w(h) = \Phi(X) - \Phi(\sqrt{14}h)$, $\kappa(t) = \frac{15+t^2}{14}$ and $\lambda(t) = 1 + \beta^2(1+t^2)$. Inserting (2), writing the constant term as $\int_{x_1}^X \varphi$, and exchanging the order of integration (Tonelli: T is an integral of a

positive function, and so are the two parts $\varphi(h)\Phi(\beta h)$ and $2\varphi(h)T(\beta h, \alpha)$ of the integrand of L , we get

$$\begin{aligned} \int_{x_1}^X \varphi(x) T\left(\frac{x}{\sqrt{14}}, \beta\right) dx &= \frac{1}{2\pi} \int_0^\beta \frac{\Phi(\sqrt{\kappa}X) - \Phi(\sqrt{\kappa}x_1)}{(1+t^2)\sqrt{\kappa}} dt, \\ \int_{x_1}^X \varphi(x) L\left(\frac{x}{\sqrt{14}}\right) dx &= (\Phi(X) - \Phi(x_1))L(h_1) + 3 \int_{h_1}^H \varphi(h)\Phi(\beta h)w(h) dh \\ &\quad - \frac{3}{\pi} \int_0^\alpha \frac{1}{1+t^2} \int_{h_1}^H \varphi(h)e^{-\beta^2(1+t^2)h^2/2}w(h) dh dt, \\ L(h_1) &= 3 \int_{-\infty}^{h_1} \varphi(h)\Phi(\beta h) dh - \frac{3}{\pi} \int_0^\alpha \frac{\Phi(\sqrt{\lambda}h_1)}{(1+t^2)\sqrt{\lambda}} dt, \end{aligned}$$

using $\int_u^v \varphi(x)e^{-\theta x^2/2}dx = \frac{\Phi(\sqrt{1+\theta}v) - \Phi(\sqrt{1+\theta}u)}{\sqrt{1+\theta}}$. The integrands are analytic near the (compact) intervals of integration; the integral over $(-\infty, h_1]$ is cut at -40 , with an error in $[0, \Phi(-40)]$. The script `e2_lower_rig.py` encloses all of them with Arb's rigorous numerical integration at 128-bit precision (square roots with branch-cut checks) and obtains $\int_{x_1}^X \gamma_q \varphi dx \in 3.16639982094540644 \times 10^{-4} \pm 7 \cdot 10^{-22}$. It also compares (2) with tensor Gauss–Hermite quadrature (difference below $1.3 \cdot 10^{-15}$ at nine points). The last statement follows from $s_{E2}^U \geq \text{val}(q) = \frac{5}{8} + J$ (Lemma 2.3(a)) and Proposition 4.2. $\square$

With Theorem 3.2 this proves Theorem A(a)–(c); part (d) is proved in §4.2. (Floating point, $\text{val}^{\text{lab}}(z \text{sgn})$ exceeds $\frac{5}{8} + \frac{3}{8}J$ by about $5.7 \cdot 10^{-9}$; see the paragraph on numerical values for E2 at the end of §1.3, and §11 for other label families.)

Remark 4.4 (few labels cannot reach s_{E2}^U). For every family (Q_z) with k label bits, $\text{val}^{\text{lab}}(Q) - \frac{5}{8} \leq (1 - 4^{-k})(s_{E2}^U - \frac{5}{8})$. Indeed, write $q_j^{(W)} = Q_{W,j}$, so that $\gamma_\nu = \mathbb{E}[\gamma_{q^{(W)}} \mid g(W) = \nu]$, and let D be the event that the three columns of W coincide. On D , $q^{(W)} = (q, q, q)$ and $\gamma_{q^{(W)}} < 0$ everywhere by Theorem 3.2. By convexity of $t \mapsto t^+$, $\gamma_\nu^+ \leq \mathbb{E}[\gamma_{q^{(W)}}^+ \mathbf{1}_{D^c} \mid g(W) = \nu]$. Moreover, by Lemma 2.3(a), $\mathbb{E}_x \gamma_q^+ = \text{val}(q) - \frac{5}{8} \leq s_{E2}^U - \frac{5}{8}$ for every mean-zero $[-1, 1]$ -valued q . Then Lemma 2.3(a) and Fubini give $\text{val}^{\text{lab}}(Q) - \frac{5}{8} = \sum_\nu \Pr[\nu] \mathbb{E}_x \gamma_\nu^+ \leq \Pr[D^c](s_{E2}^U - \frac{5}{8})$, and $\Pr[D] = 4^{-k}$. In particular one label bit yields at most $\frac{3}{4}$ of the gain of s_{E2}^U . Whether $s_{E2} = s_{E2}^U$ therefore depends on constructions with many labels.

4.2 An upper bound for s_{E2}^U : proof of Theorem A(d)

Throughout this subsection $g = \text{E2}$, $\rho = 1/\sqrt{15}$ and $c = -\frac{1}{14}$. For $t \in [-1, 1]$ let

$$\kappa(t) = \frac{1+3t}{4}, \quad \ell(t) = 2\varphi(\Phi^{-1}(\frac{1+t}{2})), \quad \sigma(t) = \sqrt{1-t^2},$$

with $\ell(\pm 1) = 0$. Let $\mathcal{E}_3$ be the set of positive semidefinite 3×3 matrices with unit diagonal, and for $w = (w_{12}, w_{13}, w_{23}) \in \mathbb{R}^3$ let $\mathbf{G}(w) = \max_{X \in \mathcal{E}_3} (w_{12}X_{12} + w_{13}X_{13} + w_{23}X_{23})$. For $m \in [-1, 1]^3$ write $\kappa_j = \kappa(m_j)$, $\ell_j = \ell(m_j)$, $\sigma_j = \sigma(m_j)$, let $K_3 = 5.15$, and

$$\begin{aligned} \Xi(m) &= \text{E2}(m) + |c| \mathbf{G}(\kappa_3 \ell_1 \ell_2, \kappa_2 \ell_1 \ell_3, \kappa_1 \ell_2 \ell_3) + c^2 (|\kappa_3| \sigma_1 \sigma_2 + |\kappa_2| \sigma_1 \sigma_3 + |\kappa_1| \sigma_2 \sigma_3) \\ &\quad + \frac{3}{4} \left(\sqrt{2} c^2 (\ell_1 \ell_2 \sigma_3 + \ell_1 \ell_3 \sigma_2 + \ell_2 \ell_3 \sigma_1) + K_3 |c|^3 \sigma_1 \sigma_2 \sigma_3 \right). \end{aligned} \tag{3}$$

Theorem 4.5. Let $\psi: [-1, 1] \rightarrow [0, \infty)$ be convex and piecewise linear, with $\psi(m_1) + \psi(m_2) + \psi(m_3) \geq \Xi(m)$ for every $m \in [-1, 1]^3$ with $\Xi(m) > 0$. Then

$$s_{E2}^U \leq \frac{5}{8} + 3 \mathbb{E} \psi(2\Phi(X/\sqrt{14}) - 1), \quad X \sim N(0, 1).$$

The piecewise linear function ψ with 401 knots computed by `numerics/e2_upper_rig.py` satisfies the hypothesis, and for it $3 \mathbb{E} \psi(2\Phi(X/\sqrt{14}) - 1) = 0.00195034886458\dots$ (Remark 4.9). Hence $s_{E2} \leq s_{E2}^U < \frac{5}{8} + 1.95035 \times 10^{-3} < 0.626951$.

The proof has three ingredients. Lemma 4.6 bounds $\gamma_q(x)$ by Ξ evaluated at the three conditional means $m_j(x) = \mathbb{E}[q_j(\mathcal{G}_j) \mid \mathcal{G}_0 = x]$; it uses the Hermite expansions of the three functions, the level-one inequality (Lemma 2.2) and the closed form of $\mathbf{G}$ (Lemma 4.7). Lemma 4.8 shows that each m_j is dominated in the convex order by the conditional mean of a halfspace, a consequence of Borell's noise stability theorem. Finally, a separable convex majorant of Ξ^+ is computed and checked with interval arithmetic.

Lemma 4.6 (pointwise bound). Let Y_1, Y_2, Y_3 be jointly Gaussian standard Gaussian vectors in $\mathbb{R}^n$ such that the n coordinates of (Y_1, Y_2, Y_3) are i.i.d. with $\mathbb{E}[Y_{i,r}Y_{j,r}] = c$ for $i \neq j$. Let $f_1, f_2, f_3: \mathbb{R}^n \rightarrow \{-1, 1\}$ be measurable and $m_j = \mathbb{E}f_j(Y_j)$. Then $\mathbb{E}E2(f_1(Y_1), f_2(Y_2), f_3(Y_3)) \leq \Xi(m_1, m_2, m_3)$.

Proof. Write $F_j = f_j(Y_j)$, $\tilde{F}_j = F_j - m_j$, $C_{ij} = \mathbb{E}[\tilde{F}_i\tilde{F}_j]$ and $T = \mathbb{E}[\tilde{F}_1\tilde{F}_2\tilde{F}_3]$. Inserting $\mathbb{E}[F_iF_j] = m_i m_j + C_{ij}$ and $\mathbb{E}[F_1F_2F_3] = m_1 m_2 m_3 + m_1 C_{23} + m_2 C_{13} + m_3 C_{12} + T$ into $E2(y) = -\frac{1}{4} + \frac{1}{4} \sum_j y_j - \frac{1}{4} \sum_{i < j} y_i y_j - \frac{3}{4} y_1 y_2 y_3$ gives

$$\mathbb{E}E2(F_1, F_2, F_3) = E2(m) - \kappa_3 C_{12} - \kappa_2 C_{13} - \kappa_1 C_{23} - \frac{3}{4} T. \quad (4)$$

Let $f_j = \sum_{d \geq 0} f_j^{\overline{d}}$ be the Hermite expansions. Then $f_j^{\overline{0}} = m_j$, $f_j^{\overline{1}}(y) = \langle a_j, y \rangle$ with $a_j = \mathbb{E}[f_j(Y)Y]$, and $\sum_{d \geq 1} \|f_j^{\overline{d}}\|_2^2 = \text{Var } f_j = \sigma_j^2$ because $f_j^2 = 1$. Lemma 2.2 applied to $b = \frac{1+f_j}{2}$, for which $\mathbb{E}b = \frac{1+m_j}{2}$ and $W_1(b) = \frac{1}{4}|a_j|^2$, gives $|a_j| \leq \ell_j$.

Pair terms. By (1) with $r_0 = c$, $C_{ij} = \sum_{d \geq 1} c^d \langle f_i^{\overline{d}}, f_j^{\overline{d}} \rangle = c \langle a_i, a_j \rangle + R_{ij}$, where $|R_{ij}| \leq c^2 \sum_{d \geq 2} \|f_i^{\overline{d}}\|_2 \|f_j^{\overline{d}}\|_2 \leq c^2 \sigma_i \sigma_j$ by Cauchy-Schwarz. Write $a_j = \ell_j v_j$ with $|v_j| \leq 1$ ($v_j = 0$ if $\ell_j = 0$) and let $u_j = v_j \oplus \sqrt{1 - |v_j|^2} e_j \in \mathbb{R}^n \oplus \mathbb{R}^3$, where e_1, e_2, e_3 is the standard basis of $\mathbb{R}^3$. The u_j are unit vectors and $\langle u_i, u_j \rangle = \langle v_i, v_j \rangle$ for $i \neq j$, so their Gram matrix lies in $\mathcal{E}_3$. As $c < 0$,

$$-\kappa_3 C_{12} - \kappa_2 C_{13} - \kappa_1 C_{23} \leq |c| \mathbf{G}(\kappa_3 \ell_1 \ell_2, \kappa_2 \ell_1 \ell_3, \kappa_1 \ell_2 \ell_3) + c^2 (|\kappa_3| \sigma_1 \sigma_2 + |\kappa_2| \sigma_1 \sigma_3 + |\kappa_1| \sigma_2 \sigma_3).$$

The triple term. Let $h_a = He_a / \sqrt{a!}$ be the normalised Hermite polynomials, $h_\alpha(y) = \prod_r h_{\alpha_r}(y_r)$ and $\alpha! = \prod_r \alpha_r!$ for $\alpha \in \mathbb{N}^n$. Then $f_j^{\overline{d}} = \sum_{|\alpha|=d} \hat{f}_j(\alpha) h_\alpha$ (a finite sum, as n is finite) and $\|f_j^{\overline{d}}\|_2^2 = \sum_{|\alpha|=d} \hat{f}_j(\alpha)^2$. If X_1, X_2, X_3 are standard Gaussians with all pairwise correlations equal to c , comparing the coefficients of $t_1^{a_1} t_2^{a_2} t_3^{a_3}$ in $\mathbb{E} \prod_j e^{t_j X_j - t_j^2/2} = e^{c(t_1 t_2 + t_1 t_3 + t_2 t_3)}$, where $e^{tx - t^2/2} = \sum_a He_a(x) t^a / a!$, gives

$$\mathbb{E}[h_{a_1}(X_1) h_{a_2}(X_2) h_{a_3}(X_3)] = \frac{\sqrt{a_1! a_2! a_3!}}{k_{12}! k_{13}! k_{23}!} c^{k_{12} + k_{13} + k_{23}}$$

if $a_1 = k_{12} + k_{13}$, $a_2 = k_{12} + k_{23}$, $a_3 = k_{13} + k_{23}$ for integers $k_{ij} \geq 0$, and 0 if there are no such integers. Fix $d = (d_1, d_2, d_3)$ with all $d_i \geq 1$ and put $E_d = \mathbb{E}[f_1^{\overline{d}_1}(Y_1) f_2^{\overline{d}_2}(Y_2) f_3^{\overline{d}_3}(Y_3)]$. As the coordinates are independent, for $|\alpha| = d_1$, $|\beta| = d_2$, $|\delta| = d_3$ the expectation $\mathbb{E}[h_\alpha(Y_1) h_\beta(Y_2) h_\delta(Y_3)]$ is the product of the one-dimensional values; it vanishes unless $\alpha = \lambda_{12} + \lambda_{13}$, $\beta = \lambda_{12} + \lambda_{23}$, $\delta = \lambda_{13} + \lambda_{23}$

for multi-indices $\lambda_{ij} \in \mathbb{N}^n$ (which are then unique), and then it equals $c^K \sqrt{\alpha! \beta! \delta!} / (\lambda_{12}! \lambda_{13}! \lambda_{23}!)$ with $K = \frac{d_1+d_2+d_3}{2}$. Hence $E_d = 0$ unless $k_{12} = \frac{d_1+d_2-d_3}{2}$, $k_{13} = \frac{d_1+d_3-d_2}{2}$ and $k_{23} = \frac{d_2+d_3-d_1}{2}$ are nonnegative integers, and in that case

$$E_d = c^K \sum_{|\lambda_{ij}|=k_{ij}} A_1(\lambda_{12}, \lambda_{13}) A_2(\lambda_{12}, \lambda_{23}) A_3(\lambda_{13}, \lambda_{23}), \quad A_j(\lambda, \lambda') = \hat{f}_j(\lambda + \lambda') \sqrt{\frac{(\lambda + \lambda')!}{\lambda! \lambda'!}}.$$

For finite matrices A, B, D and $\alpha \in \mathbb{N}^n$ we have

$$\left| \sum_{x,y,z} A(x,y) B(x,z) D(y,z) \right| \leq \|A\|_F \|B\|_F \|D\|_F, \quad \sum_{\lambda \leq \alpha, |\lambda|=k} \frac{\alpha!}{\lambda! (\alpha - \lambda)!} = \binom{|\alpha|}{k}$$

(Cauchy–Schwarz in x , then in (y, z) ; Vandermonde’s identity). Hence $\|A_1\|_F^2 = \binom{d_1}{k_{12}} \|f_1^{\neg d_1}\|_2^2$, $\|A_2\|_F^2 = \binom{d_2}{k_{12}} \|f_2^{\neg d_2}\|_2^2$ and $\|A_3\|_F^2 = \binom{d_3}{k_{13}} \|f_3^{\neg d_3}\|_2^2$. Therefore

$$|E_d| \leq M(d) |c|^K \prod_{j=1}^3 \|f_j^{\neg d_j}\|_2, \quad M(d)^2 = \binom{d_1}{k_{12}} \binom{d_2}{k_{12}} \binom{d_3}{k_{13}} = \frac{d_1! d_2! d_3!}{(k_{12}! k_{13}! k_{23}!)^2}. \quad (5)$$

Next, $T = \lim_{D \rightarrow \infty} \sum_{d_1, d_2 \leq D} \sum_{d_3 \geq 1} E_d$, where the inner sums are finite. Indeed, let P_D denote the projection onto the chaoses of degrees $1, \dots, D$. As $|\tilde{F}_j| \leq 2$, $|T - \mathbb{E}[P_D \tilde{F}_1 P_D \tilde{F}_2 \tilde{F}_3]| \leq 4 \|\tilde{F}_1 - P_D \tilde{F}_1\|_2 + 2\sigma_1 \|\tilde{F}_2 - P_D \tilde{F}_2\|_2 \rightarrow 0$. Given Y_3 we have $Y_i = cY_3 + \sqrt{1 - c^2} Z_i$ ($i = 1, 2$) with (Z_1, Z_2) independent of Y_3 , so $\mathbb{E}[f_1^{\neg d_1}(Y_1) f_2^{\neg d_2}(Y_2) | Y_3]$ is a polynomial in Y_3 of degree at most $d_1 + d_2$, and therefore $\mathbb{E}[f_1^{\neg d_1}(Y_1) f_2^{\neg d_2}(Y_2) \tilde{F}_3] = \sum_{d_3=1}^{d_1+d_2} E_d$.

Since $d_1 + d_2 + d_3 \geq 3$ is even whenever $E_d \neq 0$, every nonzero term has $K \geq 2$, and $K = 2$ only for the permutations of $(1, 1, 2)$, for which $M(d) = \sqrt{2}$; by (5), $\|f_j^{\neg 1}\|_2 = |a_j| \leq \ell_j$ and $\|f_j^{\neg 2}\|_2 \leq \sigma_j$ their contribution is at most $\sqrt{2} c^2 (\ell_1 \ell_2 \sigma_3 + \ell_1 \ell_3 \sigma_2 + \ell_2 \ell_3 \sigma_1)$. For the terms with $K \geq 3$, by (5) and Cauchy–Schwarz,

$$\sum_{K \geq 3} |E_d| \leq \left(\sum_{K \geq 3} M(d)^2 c^{2K} \right)^{1/2} \left(\sum_d \prod_j \|f_j^{\neg d_j}\|_2^2 \right)^{1/2} \leq K_3 |c|^3 \sigma_1 \sigma_2 \sigma_3,$$

where the sums run over the triples d with $d_i \geq 1$ for which the k_{ij} are nonnegative integers. Here $\sum_{K \geq 3} M(d)^2 c^{2K} = c^6 (26 + 96c^2 + 348c^4 + \dots) = 26.49903 \dots c^6 < K_3^2 c^6$: the terms with $K \leq 60$ are summed in exact rational arithmetic, and for $K > 60$ we use $M(d)^2 \leq 2^{d_1+d_2+d_3} = 4^K$ and the fact that there are at most $\frac{(K+1)(K+2)}{2}$ triples (k_{12}, k_{13}, k_{23}) with sum K , so that the tail is at most $\sum_{K > 60} \frac{(K+1)(K+2)}{2} (4c^2)^K$, which is less than twice its first term because consecutive terms have ratio $\frac{K+3}{K+1} \cdot \frac{4}{196} < 0.022$ (`e2_upper_rig.py`). In particular $\sum_d |E_d| < \infty$, so $T = \sum_d E_d$ and $-\frac{3}{4}T \leq \frac{3}{4}(\sqrt{2} c^2 (\ell_1 \ell_2 \sigma_3 + \ell_1 \ell_3 \sigma_2 + \ell_2 \ell_3 \sigma_1) + K_3 |c|^3 \sigma_1 \sigma_2 \sigma_3)$. Inserting the two bounds into (4) proves the lemma. $\square$

Lemma 4.7 (the elliptope). *Let $w \in \mathbb{R}^3$ and let $a \leq b \leq d$ be $|w_{12}|, |w_{13}|, |w_{23}|$ in increasing order. If $w_{12} w_{13} w_{23} \geq 0$ then $G(w) = a + b + d$. If $w_{12} w_{13} w_{23} < 0$ then $G(w) = b + d - a$ when $bd \geq a(b + d)$, and $G(w) = \frac{1}{2}(\frac{ab}{d} + \frac{ad}{b} + \frac{bd}{a})$ otherwise. In all cases $G(w) \leq a + b + d$, and G is convex.*

Proof. G is a maximum of linear functions, hence convex, and $G(w) \leq a + b + d$ because $|X_{ij}| \leq 1$ on $\mathcal{E}_3$. Conjugation by a diagonal sign matrix $\text{diag}(s)$ maps $\mathcal{E}_3$ onto itself and replaces w_{ij} by $s_i s_j w_{ij}$; a permutation of the indices maps $\mathcal{E}_3$ onto itself and permutes the three entries of w arbitrarily.

If $w_{12}w_{13}w_{23} \geq 0$ there are signs with $s_i s_j w_{ij} \geq 0$ for all pairs: after a permutation we may assume that $w_{23} = 0$ if some entry vanishes, and then $s_1 = 1$, $s_2 = \text{sgn } w_{12}$, $s_3 = \text{sgn } w_{13}$ (with $\text{sgn } 0 := 1$) work. Then $X = \mathbf{1}\mathbf{1}^\top$ gives $a + b + d$. If $w_{12}w_{13}w_{23} < 0$, signs can be chosen with $s_i s_j w_{ij} < 0$ for all pairs, so we may assume $w_{12} = -a$, $w_{13} = -b$, $w_{23} = -d$ with $0 < a \leq b \leq d$, and $\mathbf{G}(w) = -\min_{X \in \mathcal{E}_3} (aX_{12} + bX_{13} + dX_{23})$. Every $X \in \mathcal{E}_3$ is the Gram matrix of unit vectors u_1, u_2, u_3 , and every $x \in [-1, 1]$ is $\langle u_1, u_2 \rangle$ for some unit vectors. For fixed u_1, u_2 with $\langle u_1, u_2 \rangle = x$, $bX_{13} + dX_{23} = \langle u_3, bu_1 + du_2 \rangle \geq -|bu_1 + du_2| = -\sqrt{b^2 + d^2 + 2bdx}$, with equality for a suitable unit vector u_3 . Hence the minimum is $\min_{x \in [-1, 1]} \phi(x)$ with $\phi(x) = ax - \sqrt{b^2 + d^2 + 2bdx}$, a convex function with $\phi'(x) = 0$ exactly at $x^* = \frac{b^2 d^2 / a^2 - b^2 - d^2}{2bd}$, where $\sqrt{b^2 + d^2 + 2bdx^*} = \frac{bd}{a}$. Here $x^* \geq -1$ because $\frac{bd}{a} \geq d \geq d - b$. If $x^* \geq 1$, i.e. $bd \geq a(b + d)$, the minimum is $\phi(1) = a - b - d$; otherwise it is $\phi(x^*) = -\frac{1}{2}(\frac{ab}{d} + \frac{ad}{b} + \frac{bd}{a})$. $\square$

Lemma 4.8 (majorisation by halfspaces). *Let $0 < \rho < 1$, let $q: \mathbb{R}^n \rightarrow \{-1, 1\}$ be measurable with $\mathbb{E}q = 0$, and let $f^*(x) = (U_\rho \text{sgn})(x_1) = 2\Phi(\rho x_1 / \sqrt{1 - \rho^2}) - 1$. Then $\mathbb{E}(U_\rho q - t)^+ \leq \mathbb{E}(f^* - t)^+$ for every $t \in \mathbb{R}$, and $\mathbb{E}\psi(U_\rho q) \leq \mathbb{E}\psi(f^*)$ for every convex piecewise linear $\psi: [-1, 1] \rightarrow \mathbb{R}$.*

Proof. Let $f = U_\rho q$ and let (X, Y) be ρ -correlated standard Gaussian vectors in $\mathbb{R}^n$, so that $f(X) = \mathbb{E}[q(Y) \mid X]$ and $f^*(X) = \mathbb{E}[\text{sgn}(Y_1) \mid X]$. Since $\mathbb{E}f = \mathbb{E}f^* = 0$, both sides of the first claim equal $-t$ for $t \leq -1$, and both vanish for $t \geq 1$. For $t \in (-1, 1)$ let $A = \{f > t\}$ and $Q = \{q = 1\}$, so $\gamma_n(Q) = \frac{1}{2}$, and

$$\mathbb{E}(f - t)^+ = \mathbb{E}[(q(Y) - t)\mathbf{1}_A(X)] = 2\Pr[X \in A, Y \in Q] - (1 + t)\gamma_n(A).$$

By Borell's noise stability theorem [1] (see also [13, 4]), $\Pr[X \in A, Y \in Q] \leq \Pr[X \in H, Y \in H']$ for the parallel halfspaces $H = \{x_1 > \theta\}$ and $H' = \{y_1 > 0\}$ with $\gamma_n(H) = \gamma_n(A)$ (if $\gamma_n(A) \in \{0, 1\}$, take $H = \emptyset$ or $H = \mathbb{R}^n$, i.e. $\theta = \pm\infty$; then the inequality is trivial). The same identity with $\text{sgn}(y_1)$ in place of q gives $2\Pr[X \in H, Y \in H'] - (1 + t)\gamma_n(H) = \mathbb{E}[(f^*(X) - t)\mathbf{1}_H(X)] \leq \mathbb{E}(f^* - t)^+$, and the first claim follows. A convex piecewise linear ψ with interior knots $t_1 < \dots < t_r$ has the form $\psi(u) = \alpha + \beta u + \sum_i \lambda_i (u - t_i)^+$ with $\lambda_i \geq 0$, and $\mathbb{E}f = \mathbb{E}f^*$. $\square$

Proof of Theorem 4.5. Let $n \geq 1$ and $q_1, q_2, q_3: \mathbb{R}^n \rightarrow \{-1, 1\}$ with $\mathbb{E}q_j = 0$, and fix $x \in \mathbb{R}^n$. By Lemma 2.1, conditionally on $\mathcal{G}_0 = x$ we may write $\mathcal{G}_j = \rho x + \sqrt{1 - \rho^2} Y_j$ with (Y_1, Y_2, Y_3) as in Lemma 4.6: the conditional covariances are $1 - \rho^2$ and $-\rho^2$, and $-\rho^2 / (1 - \rho^2) = c$. Lemma 4.6 with $f_j(y) = q_j(\rho x + \sqrt{1 - \rho^2} y)$ gives $\gamma_q(x) \leq \Xi(m(x))$, where $m_j(x) = \mathbb{E}f_j(Y_j) = (U_\rho q_j)(x)$. If $\Xi(m(x)) \leq 0$ then $\gamma_q(x)^+ = 0$; otherwise $\gamma_q(x)^+ \leq \Xi(m(x)) \leq \sum_j \psi(m_j(x))$. As $\psi \geq 0$, in both cases $\gamma_q(x)^+ \leq \sum_j \psi(m_j(x))$. By Lemma 2.3(a) (with $\mu = -\frac{1}{4}$) and Lemma 4.8,

$$\text{val}(q) = \frac{5}{8} + \mathbb{E}_x \gamma_q^+ \leq \frac{5}{8} + \sum_{j=1}^3 \mathbb{E} \psi(U_\rho q_j) \leq \frac{5}{8} + 3 \mathbb{E} \psi(2\Phi(x_1 / \sqrt{14}) - 1),$$

because $\rho / \sqrt{1 - \rho^2} = 1 / \sqrt{14}$. Taking the supremum over n and q proves the inequality; the numerical statements are verified in Remark 4.9. $\square$

Remark 4.9 (the computer-assisted part). `python3 e2.upper_rig.py 400` (python-flint 0.9 for Arb ball arithmetic; 84 seconds) constructs ψ and verifies its properties. (i) *Knots*. The values $\Phi^{-1}(\frac{1+u}{2})$, $u = -1 + \frac{k}{200}$ ($0 < k < 400$), rounded to multiples of 2^{-30} , give $z_1 < \dots < z_{399}$ with $z_{200} = 0$ and $|z_a| < 2.81$; the knots are $u_0 = -1$, $u_a = 2\Phi(z_a) - 1$ and $u_{400} = 1$. At the knots $\ell(u_a) = 2\varphi(z_a)$ and $\sigma(u_a) = \sqrt{1 - u_a^2}$; these are enclosed in Arb balls (200 bits) and converted to double precision intervals with outward rounding. ψ interpolates values $\psi(u_a) =$

$y_a \geq 0$ that are double precision numbers (outputs/e2_upper_psi_400.txt); they come from a linear programme (minimise a discretisation of $3\mathbb{E}\psi(U)$ subject to the vertex inequalities of (iii) and convexity), increased by $2 \cdot 10^{-9} + 10^{-6}u_a^2$. (ii) *Convexity*. The 399 numbers $(y_{a+2} - y_{a+1})(u_{a+1} - u_a) - (y_{a+1} - y_a)(u_{a+2} - u_{a+1})$, which are the slope increments of ψ multiplied by $(u_{a+1} - u_a)(u_{a+2} - u_{a+1})$, are positive in ball arithmetic with the exact knots (the smallest lower bound is $2.5 \cdot 10^{-13}$; the smallest slope increment is $1.0 \cdot 10^{-8}$). (iii) *The majorant property*. The cells $B = [u_a, u_{a+1}] \times [u_b, u_{b+1}] \times [u_c, u_{c+1}]$ cover $[-1, 1]^3$, and by the symmetry of Ξ and of $\sum_j \psi(m_j)$ it suffices to treat $a \leq b \leq c$ (10,746,800 cells). (Ξ is symmetric: permuting m permutes the three arguments of G and the terms of the other sums in (3), and by Lemma 4.7 $G(w)$ depends only on the multiset $\{|w_{12}|, |w_{13}|, |w_{23}|\}$ and the sign of $w_{12}w_{13}w_{23}$.) On B the function $\sum_j \psi(m_j) - E2(m)$ is multilinear, so its minimum over B is attained at a vertex. On the other hand $\Xi - E2 \leq \text{Corr}(B)$ on B , where $\text{Corr}(B)$ is computed from (3) as follows. As 0 is a knot, ℓ and σ are monotone on every knot interval, so their ranges there are spanned by their values at the two endpoints; κ is affine. This gives a box containing all $(\kappa_3\ell_1\ell_2, \kappa_2\ell_1\ell_3, \kappa_1\ell_2\ell_3)$, $m \in B$, and since G is convex it is bounded on the box by its maximum over the 8 corners, evaluated with the closed form of Lemma 4.7; the other terms are bounded by products of upper endpoints. If $E2(v) + \text{Corr}(B) \leq 0$ at all vertices v of B , then $\Xi \leq 0$ on B and nothing is required. For each of the remaining 2,884,855 cells the script checks $\psi(v_1) + \psi(v_2) + \psi(v_3) \geq E2(v) + \text{Corr}(B)$ at the 8 vertices, which implies $\sum_j \psi(m_j) \geq \Xi(m)$ on B ; the smallest slack is $5.6 \cdot 10^{-7}$. (iv) *Rounding*. The knot data are enclosed with outward rounding as in (i); the inequalities of (iii) are then evaluated in double precision with rounding to nearest, and the rounding errors are absorbed by explicit margins: $\text{Corr}(B)$ is multiplied by $1 + 10^{-12}$ and increased by 10^{-14} , $E2(v)$ is increased by 10^{-14} , and every comparison is required to hold with slack above 10^{-13} . All terms of $\text{Corr}(B)$ are nonnegative; the only subtraction, $b + d - a$ in Lemma 4.7, satisfies $b + d - a \geq \frac{1}{3}(a + b + d)$; and G is 1-Lipschitz for the ℓ^1 norm with $G(w) \geq b + d - a \geq \frac{1}{3}\|w\|_1$ (take $X = ss^T$ for suitable signs s), so relative errors of its arguments are amplified at most threefold. Counting the roundings along each term (at most 25 relative errors of size 2^{-53} , including this amplification), the relative error of the computed $\text{Corr}(B)$ is below $30 \cdot 2^{-53} < 4 \cdot 10^{-15}$, far below 10^{-12} . The absolute errors are below 10^{-14} : the enclosure of κ is widened by one ulp only, which near its zero $u = -\frac{1}{3}$ does not cover the rounding error of $(1 + 3u)/4$, but that error is below $2 \cdot 10^{-16}$ and changes $\text{Corr}(B)$ by less than 10^{-16} ; the evaluation of $E2(v)$ (about a dozen operations on numbers of modulus at most 3) has error below $4 \cdot 10^{-15}$; and the double precision vertices differ from the exact knots by at most $2.3 \cdot 10^{-16}$ per coordinate, which changes $E2$ by less than $7 \cdot 10^{-16}$ by (M2). The slacks (sums of four numbers of modulus at most 3) are computed with error below $2 \cdot 10^{-15}$, far below 10^{-13} . (v) *The expectation*. On $[u_a, u_{a+1}]$, $\psi(u) = \alpha_a + \beta_a u$, and $U = 2\Phi(X/\sqrt{14}) - 1 \in [u_a, u_{a+1}]$ iff $X \in [\sqrt{14}z_a, \sqrt{14}z_{a+1}]$ ($z_0 = -\infty, z_{400} = \infty$). So $\mathbb{E}\psi(U) = \sum_a (\alpha_a P_a + \beta_a M_a)$ with $P_a = \Phi(\sqrt{14}z_{a+1}) - \Phi(\sqrt{14}z_a)$ and $M_a = \int_{\sqrt{14}z_a}^{\sqrt{14}z_{a+1}} (2\Phi(s/\sqrt{14}) - 1)\varphi(s) ds$, evaluated with Arb's rigorous numerical integration; on the two outer intervals the integral is cut at $|s| = 40$, with error at most $\Phi(-40)$. The result is $3\mathbb{E}\psi(U) \in [0.0019503488645821, 0.0019503488645822]$. With 100, 200 and 300 instead of 400 subintervals the same procedure gives $s_{E2}^U \leq 0.6270294, 0.6269737$ and 0.6269565 .

Remark 4.10 (how sharp is Theorem A(d)?). The bound gives $s_{E2}^U - \frac{5}{8} < 1.951 \cdot 10^{-3}$, whereas $(\text{sgn}, \text{sgn}, -\text{sgn})$ gives $J = 3.1664 \cdot 10^{-4}$ (Proposition 4.3). For this configuration $m_1(x) = m_2(x) = f^*(x)$ and $m_3(x) = -f^*(x)$, so Lemma 4.8 is an identity, and the loss has two sources (e2_upper_sanity.py, floating point): along the curve $m(x)$, $\mathbb{E}_x \Xi(m(x))^+ = 1.211 \cdot 10^{-3}$, a factor 3.8 above $\mathbb{E}\gamma^+$, and $\mathbb{E}_x \sum_j \psi(m_j(x)) = 1.950 \cdot 10^{-3}$, a further factor 1.6. The second loss comes from $x < 0$, where $m(x) = (-t, -t, t)$ with $t > 0$ and $\Xi(m(x)) < 0$, but $\psi(\pm t) > 0$ is forced by other points of the cube. More general majorants of the same kind do not reduce it in our

experiments. Lemma 4.8 holds for every measurable $h: \mathbb{R}^n \rightarrow [-1, 1]$ with $\mathbb{E}h = 0$ (in its proof, $\mathbb{E}[h(Y)\mathbf{1}_A(X)]$ is maximised, by the bathtub principle, by some $h = 2 \cdot \mathbf{1}_Q - 1$ with $\gamma_n(Q) = \frac{1}{2}$), and $U_\rho(\sum_j \theta_j q_j) = \langle \theta, m \rangle$; so $\sum_j \psi(m_j)$ may be replaced by $\sum_\theta \psi_\theta(\langle \theta, m \rangle)$ with convex ψ_θ and finitely many directions $\|\theta\|_1 \leq 1$. With all vectors with entries $\pm \frac{1}{3}$, all vectors with two entries $\pm \frac{1}{2}$ and one entry 0, and the permutations of $(\frac{1}{2}, \pm \frac{1}{4}, \pm \frac{1}{4})$ (one ψ per permutation class), the discretised linear programme has the same value as the separable one up to the solver tolerance (`e2_upper_ridge.py`, floating point, 61 knots). The first loss is not a weakness of Lemma 4.6 but of every bound in terms of the conditional means. In the setting of that lemma with $n = 1$, the aligned functions $f_1 = f_2 = \text{sgn}(y - \theta)$, $f_3 = \text{sgn}(y - \theta')$ with $m = (t, t, -t)$ come close to Ξ : for $t = 0.3, 0.45, 0.6, 0.75, 0.9$ the gap $\Xi(m) - \mathbb{E}E2(F)$ decreases from 0.013 to 0.002. The mean of the positive part of their value along $m(x)$ is $1.014 \cdot 10^{-3}$. Hence every bound B with $\mathbb{E}E2(f_1(Y_1), f_2(Y_2), f_3(Y_3)) \leq B(m_1, m_2, m_3)$ for all Boolean f_j with given means (as in Lemma 4.6) has $\mathbb{E}_x B(m(x))^+ \geq 1.014 \cdot 10^{-3}$ along the curve $m(x)$ of $(\text{sgn}, \text{sgn}, -\text{sgn})$ (floating point); so no argument that bounds $\gamma_q(x)$ through such a function of $m(x)$ alone can prove $s_{E2}^U < 0.62601$. Going further requires using that the functions f_j for different x are restrictions of the same q_j (for the aligned choice at every x , m_3 would be increasing in x , while $m_3 = -f^*$ is decreasing). For random halfspaces in $\mathbb{R}^n$, $n \leq 3$, and random finite unions of intervals in $\mathbb{R}$ the gap $\Xi(m) - \mathbb{E}E2(F)$ is always positive in our tests, with smallest values $6.0 \cdot 10^{-4}$ and $5.0 \cdot 10^{-3}$.

5 The lower bound $s_g \geq s_g^{\text{lab}}$

Theorem 5.1. *Let $g: \{-1, 1\}^m \rightarrow \{-1, 1\}$ depend on at least two coordinates, with $\mu = \mathbb{E}g \neq 0$. Then $s_g \geq s_g^{\text{lab}}$.*

Fix k, n and a family (Q_z) as in Definition 1.5, and measurable “decoders” $P_\nu: \mathbb{R}^n \rightarrow [-1, 1]$, $\nu \in \{-1, 1\}^k$. Put

$$V(P, Q) = \mathbb{E}_W \mathbb{E}_{g \sim \mathcal{N}_g^n} \left[P_{g(W)}(\mathcal{G}_0) g(Q_{W_1}(\mathcal{G}_1), \dots, Q_{W_m}(\mathcal{G}_m)) \right] = \sum_\nu \Pr[g(W) = \nu] \mathbb{E}_x [P_\nu(x) \gamma_\nu(x)].$$

With $P_\nu = \text{sgn } \gamma_\nu$ we get $\frac{1+V(P, Q)}{2} = \text{val}^{\text{lab}}(Q)$. It therefore suffices to prove:

Proposition 5.2. *For all (P, Q) as above and every $\eta > 0$ there is a sequence of Boolean functions f_N ($N \rightarrow \infty$) with $\max_{|S| \leq L} |f_N(S)| \rightarrow 0$ for every fixed L and $\liminf_N \Pr[f_N \circ g = g \circ f_N] \geq \frac{1+V(P, Q)}{2} - \eta$.*

By Remark 1.2(a) the proposition gives $s_g \geq \frac{1+V(P, Q)}{2} - \eta$ for every $\eta > 0$, hence $s_g \geq \text{val}^{\text{lab}}(Q)$ (with $P_\nu = \text{sgn } \gamma_\nu$), and Theorem 5.1 follows by taking the supremum over Q .

Step 0: continuity. By (M2), $|g(y) - g(y')| \leq \sum_j |y_j - y'_j|$ on $[-1, 1]^m$; since each $\mathcal{G}_j$ is $N(0, I_n)$ and independent of W ,

$$|V(P, Q) - V(P', Q')| \leq \max_\nu \|P_\nu - P'_\nu\|_{L^1(\gamma_n)} + m \max_z \|Q_z - Q'_z\|_{L^1(\gamma_n)}.$$

Continuous functions are dense in $L^1(\gamma_n)$, and truncation to $[-1, 1]$ does not increase the distance to a $[-1, 1]$ -valued function. If Q'_z is continuous, $[-1, 1]$ -valued and $\|Q'_z - Q_z\|_1 \leq \eta'$, then $Q''_z = (Q'_z - \mathbb{E}Q'_z)/(1 + |\mathbb{E}Q'_z|)$ is continuous, $[-1, 1]$ -valued, has mean 0, and $\|Q''_z - Q_z\|_1 \leq 3\eta'$ (as $|\mathbb{E}Q'_z| \leq \eta'$). Hence we may and do assume that all P_ν, Q_z are continuous.

Step 1: the construction. Assume $\mu < 0$ (for $\mu > 0$ reverse the inequalities defining $\mathcal{A}, \mathcal{B}$). Let $K = k + nN$; we index the coordinates of $\{-1, 1\}^K$ by $[k]$ (labels) and by pairs $(b, i) \in [n] \times [N]$ (block b). For $v \in \{-1, 1\}^{[n] \times [N]}$ put $\sigma(v) = \frac{1}{nN} \sum_{b,i} v_{b,i}$,

$$\bar{v}_b = \frac{1}{\sqrt{N}} \sum_i v_{b,i}, \quad \tilde{v}_b = \frac{\sum_i v_{b,i} - N\mu}{\sqrt{N(1-\mu^2)}} \quad (b \in [n]),$$

$\mathcal{A} = \{v : \sigma(v) \geq \mu/2\}$, $\mathcal{B} = \{v : \sigma(v) < \mu/2\}$, and define $f^* : \{-1, 1\}^k \times \{-1, 1\}^{[n] \times [N]} \rightarrow [-1, 1]$ by

$$f^*(z, v) = Q_z(\bar{v}) \text{ if } v \in \mathcal{A}, \quad f^*(z, v) = P_z(\tilde{v}) \text{ if } v \in \mathcal{B}.$$

Step 2: the value of f^* . Let Z be uniform in $\{-1, 1\}^{K \times m}$, $W = (Z_{ij})_{i \in [k]}$ its label rows, and for $j \in [m]$ let $v^{(j)} = (Z_{(b,i),j})_{b,i}$ and $v^{(0)} = (g(Z_{(b,i),\cdot}))_{b,i}$. Then $\text{col}_j Z = (W_j, v^{(j)})$ and $g^K(Z) = (g(W), v^{(0)})$, and W is independent of $(v^{(0)}, \dots, v^{(m)})$. The coordinates of $v^{(0)}$ are i.i.d. with mean μ , those of $v^{(j)}$ i.i.d. uniform; by Hoeffding's inequality the event $E = \{v^{(0)} \in \mathcal{B}, v^{(1)}, \dots, v^{(m)} \in \mathcal{A}\}$ has $\Pr[E^c] \leq (m+1)e^{-nN\mu^2/8}$. On E ,

$$\begin{aligned} f^*(g^K(Z)) g(f^*(\text{col}_1 Z), \dots, f^*(\text{col}_m Z)) &= P_{g(W)}(\tilde{v}^{(0)}) g(Q_{W_1}(\bar{v}^{(1)}), \dots, Q_{W_m}(\bar{v}^{(m)})) \\ &=: \Psi_W(\tilde{v}^{(0)}, \bar{v}^{(1)}, \dots, \bar{v}^{(m)}). \end{aligned}$$

For each block b , $(\tilde{v}_b^{(0)}, \bar{v}_b^{(1)}, \dots, \bar{v}_b^{(m)}) = N^{-1/2} \sum_{i \leq N} \xi_{b,i}$ where

$$\xi_{b,i} = \left(\frac{g(Z_{(b,i),\cdot}) - \mu}{\sqrt{1-\mu^2}}, Z_{(b,i),1}, \dots, Z_{(b,i),m} \right)$$

are i.i.d. centred vectors whose covariance is that of $\mathcal{N}_g$ (CFMMS Lemma 5.10: $\mathbb{E}[\frac{g(x)-\mu}{\sqrt{1-\mu^2}}x_j] = \rho_j$), and different blocks are independent. By the multivariate CLT,

$$(\tilde{v}^{(0)}, \bar{v}^{(1)}, \dots, \bar{v}^{(m)}) \rightarrow (\mathcal{G}_0, \dots, \mathcal{G}_m) \sim \mathcal{N}_g^n \text{ in distribution.}$$

For each fixed W , Ψ_W is bounded by 1 and continuous, so $\mathbb{E}\Psi_W(\tilde{v}^{(0)}, \dots) \rightarrow \mathbb{E}\Psi_W(\mathcal{G}_0, \dots, \mathcal{G}_m)$; since the left-hand side of the last display is bounded by 1, W is independent of the v 's, and $\Pr[E^c] \rightarrow 0$,

$$V_N := \mathbb{E}_Z \left[f^*(g^K(Z)) g(f^*(\text{col}_1 Z), \dots, f^*(\text{col}_m Z)) \right] \xrightarrow{N \rightarrow \infty} V(P, Q). \quad (6)$$

Step 3: f^* is pseudorandom. Write $f^* = \Upsilon + R$ with $\Upsilon(z, v) = Q_z(\bar{v})$ and $R = \mathbf{1}_{\mathcal{B}}(v)(P_z(\tilde{v}) - Q_z(\bar{v}))$. Under the uniform measure, $\Pr[v \in \mathcal{B}] \leq e^{-nN\mu^2/8}$, so $|\hat{R}(S)| \leq 2e^{-nN\mu^2/8}$ for all S . For $S = S_1 \sqcup S_2$ with $S_1 \subseteq [k]$ and $S_2 \subseteq [n] \times [N]$, $\hat{\Upsilon}(S) = \mathbb{E}_z[\chi_{S_1}(z)\hat{\phi}_z(S_2)]$ where $\phi_z(v) = Q_z(\bar{v})$. If $S_2 = \emptyset$ then $|\hat{\phi}_z(\emptyset)| = |\mathbb{E}_v Q_z(\bar{v})| \rightarrow |\mathbb{E} Q_z(\mathcal{G})| = 0$ by the CLT. If $1 \leq |S_2| \leq N-1$: ϕ_z is invariant under permutations of the coordinates inside each block, so all sets in the orbit of S_2 under $\prod_b \text{Sym}(N)$ have the same coefficient; the orbit has $\prod_b \binom{N}{L_b} \geq N$ elements ($L_b = |S_2 \cap (\{b\} \times [N])|$, some $L_b \in [1, N-1]$), so Parseval gives $N\hat{\phi}_z(S_2)^2 \leq \mathbb{E}\phi_z^2 \leq 1$. Altogether, uniformly in the degree up to $N-1$,

$$\epsilon_N := \max_{|S| \leq N-1} |\widehat{f^*}(S)| \leq \max_z \left(\max_z |\mathbb{E}_v Q_z(\bar{v})|, N^{-1/2} \right) + 2e^{-nN\mu^2/8} \xrightarrow{N \rightarrow \infty} 0.$$

Step 4: rounding to a Boolean function. Let f be random with independent values $f(u) \in \{-1, 1\}$, $\mathbb{E}f(u) = f^*(u)$ ($u \in \{-1, 1\}^K$). For a fixed Z whose $m+1$ points $g^K(Z), \text{col}_1 Z, \dots, \text{col}_m Z$ are distinct, independence gives $\mathbb{E}_f[f(g^K Z) \prod_{j \in S} f(\text{col}_j Z)] = f^*(g^K Z) \prod_{j \in S} f^*(\text{col}_j Z)$. Two columns coincide with probability 2^{-K} , and $\Pr[g^K(Z) = \text{col}_j Z] = p_j^K$ with $p_j = \Pr_x[g(x) = x_j] = \frac{1+\hat{g}(\{j\})}{2} < 1$ (g is not the dictator x_j). Expanding g in the Fourier basis, $|\mathbb{E}_f V(f) - V_N| \leq 2\|\hat{g}\|_1 \left(\binom{m}{2} 2^{-K} + \sum_j p_j^K\right) \rightarrow 0$, where $V(f) = \mathbb{E}_Z[(f \circ g^K)(g \circ f^m)]$, so $\mathbb{E}_f \Pr[f \circ g = g \circ f] = \frac{1+\mathbb{E}_f V(f)}{2} \rightarrow \frac{1+V(P,Q)}{2}$. Moreover $\hat{f}(S) - \hat{f}^*(S) = 2^{-K} \sum_u (f(u) - f^*(u)) \chi_S(u)$ is a sum of 2^K independent centred terms of range $\leq 2^{1-K}$, so by Hoeffding $\Pr[|\hat{f}(S) - \hat{f}^*(S)| \geq 2^{-K/4}] \leq 2 \exp(-2^{K/2}/2)$. A union bound over the at most $(K+1)^{N-1}$ sets with $|S| \leq N-1$ shows that the event $G_N = \{\max_{|S| \leq N-1} |\hat{f}(S) - \hat{f}^*(S)| < 2^{-K/4}\}$ has $\Pr[G_N^c] \leq 2(K+1)^{N-1} e^{-2^{K/2}/2} \rightarrow 0$, because $K \geq N$. Since agreement probabilities lie in $[0, 1]$, some $f_N \in G_N$ has $\Pr[f_N \circ g = g \circ f_N] \geq \mathbb{E}_f \Pr[f \circ g = g \circ f] - \Pr[G_N^c]$. This defines one sequence (f_N) , and for every fixed L and $N > L$ we have $\max_{|S| \leq L} |\hat{f}_N(S)| \leq \epsilon_N + 2^{-K/4} \rightarrow 0$. Its agreement tends to $\frac{1+V(P,Q)}{2}$, where (P, Q) are the continuous approximations of Step 0, which lie within η of the original value; this proves Proposition 5.2. $\square$

6 The upper bound $s_g \leq s_g^{\text{lab}}$

Theorem 6.1. *Let g depend on all coordinates, $g \neq \pm \prod_j x_j$, $\mu = \mathbb{E}g \neq 0$. Then $s_g \leq s_g^{\text{lab}}$.*

The proof follows CFMMS's proof of their Theorem 5.3 with two changes: (1) we do not take a supremum separately for each restriction of the influential coordinates, but keep track of the joint law of the column labels and the output label; (2) to make the Gaussian test function depend only on the label, all m column functions must receive the *same* noise rate. CFMMS Theorem 5.12 (noise resilience of connected distributions) only asserts that some rates $\gamma_1, \dots, \gamma_k > 0$ work; its proof chooses them one after another (γ_i after degree cut-offs that depend on $\gamma_1, \dots, \gamma_{i-1}$), so the statement gives no control over the rates, and in particular no way to make them equal. We need equal rates on the m column parties, so we replace it by a three-party version (Lemma 6.7) in which the parties $x_{<m}$ and x_m receive the same rate.

6.1 Tools

Theorem 6.2 (Jones' regularity lemma; CFMMS Thm. 2.9). *For all $d \in \mathbb{N}$, $\tau, \epsilon > 0$ there is L such that for every $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ there is $\Lambda \subseteq [n]$, $|\Lambda| \leq L$, such that for each of the two product measures π_0 (uniform) and $\pi_\mu: \Pr_{z \sim \pi}[f_\Lambda \rightarrow z \text{ is } (d, \tau)\text{-regular w.r.t. } \pi] \geq 1 - \epsilon$. ((d, τ) -regular: $\inf_i^{\leq d} \leq \tau$ for all i .)*

Theorem 6.3 (Invariance; CFMMS Thm. 5.9, from Mossel, GAFA 2010, Thms. 4.1–4.2). *Let $\mathcal{D}$ be a distribution on $\{-1, 1\}^{k'}$ with coordinate means $\mu_1, \dots, \mu_{k'}$ and let $\mathcal{G}$ be Gaussian with the same mean vector and covariance. For all $\epsilon, \gamma > 0$ there are $d \in \mathbb{N}$, $\tau > 0$ such that for all $f_1, \dots, f_{k'}: \{-1, 1\}^n \rightarrow \{-1, 1\}$ that are (d, τ) -regular (w.r.t. π_{μ_ℓ}) and all $\gamma_1, \dots, \gamma_{k'} \geq \gamma$,*

$$\left| \mathbb{E}_{x \sim \mathcal{D}^n} \prod_{\ell} (T_{1-\gamma_\ell}^{\mu_\ell} f_\ell)(x_\ell) - \mathbb{E}_{\mathcal{G}^n} \prod_{\ell} \text{clip}(T_{1-\gamma_\ell}^{\mu_\ell} f_\ell)(\mathcal{G}_\ell) \right| < \epsilon, \quad \left| \mathbb{E} \text{clip}(T_{1-\gamma_\ell}^{\mu_\ell} f_\ell)(\mathcal{G}_\ell) - \mathbb{E} f_\ell \right| < \epsilon.$$

Lemma 6.4 (the two cases of CFMMS Lemma 5.13 that we use). *Under the hypotheses of Theorem 6.1, after permuting the coordinates of g , the following two bipartite graphs are connected*

(x uniform; vertices are the values that occur, edges join pairs of values that occur jointly): $G^{(1)}$ between the values of $x_{<m} = (x_1, \dots, x_{m-1})$ and the values of $(g(x), x_m)$, and $G^{(2)}$ between the values of $(x_{<m}, g(x))$ and the values of x_m .

Proof. If for every j we had $g(x) = x_j g'_j(x_{-j})$ for some g'_j , every set in the Fourier support of g would contain every j , so $g = \pm \prod_j x_j$. So we may order the coordinates such that, writing $x = (a, t)$ with $a \in \{-1, 1\}^{m-1}$ and $t = x_m$, some a_0 has $g(a_0, 1) = g(a_0, -1)$; as g depends on x_m , some a_1 has $g(a_1, 1) \neq g(a_1, -1)$.

$G^{(2)}$: each left vertex $(a, g(a, t))$ is adjacent to $t \in \{-1, 1\}$, and the left vertex $(a_0, g(a_0, 1))$ is adjacent to both 1 and -1 ; hence $G^{(2)}$ is connected.

$G^{(1)}$: each a is adjacent to exactly the two vertices $(g(a, 1), 1)$ and $(g(a, -1), -1)$, so $G^{(1)}$ is connected iff the graph H on the right-hand vertices with one edge $\{(g(a, 1), 1), (g(a, -1), -1)\}$ for each a is connected. If $g(\cdot, 1)$ or $g(\cdot, -1)$ is constant, all edges of H contain the same vertex and H is connected. Otherwise all four vertices $(\pm 1, \pm 1)$ occur, and H is a subgraph of the complete bipartite graph $K_{2,2}$ between $\{(\pm 1, 1)\}$ and $\{(\pm 1, -1)\}$ without isolated vertices. The edges with $g(a, 1) = g(a, -1)$ lie in one perfect matching of $K_{2,2}$ and those with $g(a, 1) \neq g(a, -1)$ in the other. H contains edges of both kinds (from a_0 and a_1); two such edges share a vertex and form a path through three vertices, which contains both vertices of one side. The fourth vertex has an edge, which ends on that side, so H is connected. $\square$

Fix such an ordering. On the single-coordinate space $\Omega = \{-1, 1\}^m$ (uniform) consider the three “parties” $A(x) = x_{<m}$, $B(x) = g(x)$, $C(x) = x_m$. For two maps L, R on a finite probability space with positive atoms, let $\rho(L, R) = \sup\{\mathbb{E}[\phi(L)\psi(R)] : \mathbb{E}\phi = \mathbb{E}\psi = 0, \mathbb{E}\phi^2 = \mathbb{E}\psi^2 = 1\}$ (maximal correlation).

Lemma 6.5. (a) If the bipartite graph of (L, R) is connected then $\rho(L, R) < 1$. (b) On the n -fold product, let $\mathcal{T}\psi = \mathbb{E}[\psi(R^{(n)}) \mid L^{(n)}]$ for $\psi \in L^2$ of $R^{(n)} = (R(\omega_1), \dots, R(\omega_n))$. If ψ has Efron–Stein decomposition $\psi = \sum_S \psi_S$ (w.r.t. the product of the R -marginals) then $\mathcal{T}\psi_S$ is the S -component of $\mathcal{T}\psi$ (w.r.t. the L -marginals) and $\|\mathcal{T}\psi_S\|_2 \leq \rho^{|S|} \|\psi_S\|_2$, $\rho = \rho(L, R)$. In particular $\|\mathcal{T}\psi\|_2 \leq \rho^{D+1} \|\psi\|_2$ if $\psi_S = 0$ for all $|S| \leq D$.

Proof. (a) The supremum is over a compact set of a finite-dimensional space, so it is attained; if it equals 1 then $\phi(L) = \psi(R)$ a.s., i.e. $\phi(l) = \psi(r)$ for every edge (l, r) ; connectivity forces ϕ, ψ constant, contradicting $\mathbb{E}\phi = 0, \mathbb{E}\phi^2 = 1$. (b) This is Mossel (GAFA 2010) Props. 2.11–2.12: $\mathcal{T} = \bigotimes_{i \leq n} \mathcal{T}^{(1)}$, where the single-coordinate operator $\mathcal{T}^{(1)}$ maps constants to constants and mean-zero functions to mean-zero functions with norm $\leq \rho$; the S -component space is $\bigotimes_{i \in S} (\text{mean zero}) \otimes \bigotimes_{i \notin S} (\text{constants})$. $\square$

Lemma 6.6 (hypercontractivity; O’Donnell Ch. 10, as used in CFMMS Lemma B.1). Let (Ω', π') be a finite probability space with minimal atom λ , and $M = 1/\lambda$. If $\phi: \Omega'^n \rightarrow \mathbb{R}$ has Efron–Stein degree $\leq e$ then $\|\phi\|_q \leq (\sqrt{q-1} M^{1/2-1/q})^e \|\phi\|_2$ for $q \geq 2$.

We also use: for ψ_1 a function of party X with all ES components of degree $\leq e_1$ and ψ_2 a function of party Y with all ES components of degree $> e_2$ (each w.r.t. its own marginal), the product $\psi_1(X)\psi_2(Y)$, viewed on the joint product space of (X, Y) , has no ES components of degree $\leq e_2 - e_1$. (For $|S| \leq e_1, |S'| > e_2$ and $|V| \leq e_2 - e_1$ pick $i \in S' \setminus (S \cup V)$; averaging $\psi_{1,S}\psi_{2,S'}$ over coordinate i gives 0, hence all conditional expectations onto coordinates in V , and so the V -component, vanish. A function of X alone has the same ES decomposition on the joint space as on its own space.)

6.2 A three-party noise lemma

Noise operators: $T_{1-\gamma}$ is the usual noise operator on $\{-1, 1\}^n$ (uniform), $T_{1-\gamma}^\mu$ its μ -biased analogue, and on party A , whose values lie in $(\{-1, 1\}^{m-1})^n \cong \{-1, 1\}^{(m-1)n}$, $T_{1-\gamma}^A = T_{1-\gamma}^{\otimes(m-1)}$ is the bitwise noise operator (each of the $(m-1)n$ bits is re-randomised with probability γ). Note $T_{1-\gamma}^A \prod_{j < m} f_j(a \cdot j) = \prod_{j < m} (T_{1-\gamma} f_j)(a \cdot j)$.

Lemma 6.7. *Under the hypotheses of Theorem 6.1 (ordering of Lemma 6.4), for every $\epsilon > 0$ there are $\gamma, \gamma_0 \in (0, 1)$, depending only on g and ϵ , such that for all n and all $F: \{-1, 1\}^{(m-1)n} \rightarrow [-1, 1]$, $f_0, f_m: \{-1, 1\}^n \rightarrow [-1, 1]$,*

$$\left| \mathbb{E}[F(A)f_0(B)f_m(C)] - \mathbb{E}[(T_{1-\gamma}^A F)(A)(T_{1-\gamma_0}^\mu f_0)(B)(T_{1-\gamma} f_m)(C)] \right| \leq \epsilon,$$

where $(A, B, C) = (A(x^{(i)}), B(x^{(i)}), C(x^{(i)}))_{i \leq n}$ with $x^{(1)}, \dots, x^{(n)}$ i.i.d. uniform in $\{-1, 1\}^m$.

Proof. Let $r_1 = \max(\frac{1}{2}, \rho(A, (B, C)))$ and $r_2 = \max(\frac{1}{2}, \rho((A, B), C))$ (maximal correlations; not to be confused with the level-one quantities ρ_j); these lie in $(0, 1)$ by Lemmas 6.4 and 6.5(a), and Lemma 6.5(b) holds with any upper bound on the maximal correlation. Let $K_2 = \max_{e \geq 0} e r_2^e$, $M_A = 2^{m-1}$, $M_B = 2/(1 - |\mu|)$. Write $T = T_{1-\gamma}^A$ or $T_{1-\gamma}$ as appropriate and $T_0 = T_{1-\gamma_0}^\mu$. We pass from $E_0 = \mathbb{E}[F f_0 f_m]$ to $E_1 = \mathbb{E}[T F f_0 f_m]$, $E_2 = \mathbb{E}[T F f_0 T f_m]$, $E_3 = \mathbb{E}[T F T_0 f_0 T f_m]$.

Step 1 (A). Let $G = (I - T)F$ and let $\mathcal{T}_1$ be conditional expectation from functions of A to functions of (B, C) . Since $|f_0 f_m| \leq 1$, $|E_0 - E_1| = |\mathbb{E}[(\mathcal{T}_1 G)(B, C) f_0(B) f_m(C)]| \leq \|\mathcal{T}_1 G\|_2$. Split $G = G^{\leq D} + G^{> D}$ (ES decomposition on the A -space). Then $\|\mathcal{T}_1 G^{> D}\|_2 \leq r_1^{D+1} \|G\|_2 \leq r_1^{D+1}$ (Lemma 6.5(b); $\|G\|_2 \leq \|F\|_2 \leq 1$ as $I - T$ has spectrum in $[0, 1]$). The ES component of degree $|S|$ on the A -space is spanned by characters of at most $(m-1)|S|$ bits, on which $I - T$ acts by $1 - (1 - \gamma)^{\# \text{bits}} \leq (m-1)|S|\gamma$; hence $\|\mathcal{T}_1 G^{\leq D}\|_2 \leq \|G^{\leq D}\|_2 \leq (m-1)D\gamma$. So $|E_0 - E_1| \leq r_1^{D+1} + (m-1)D\gamma$.

Step 2 (C). Let $\mathcal{T}_2$ map functions of C to functions of (A, B) . Since $|T F \cdot f_0| \leq 1$, $|E_1 - E_2| \leq \|\mathcal{T}_2(I - T)f_m\|_2$, and by Lemma 6.5(b) and orthogonality of ES components, $\|\mathcal{T}_2(I - T)f_m\|_2^2 \leq \sum_S r_2^{2|S|} (1 - (1 - \gamma)^{|S|})^2 \|(f_m)_S\|^2 \leq \max_e (r_2^e e \gamma)^2$. So $|E_1 - E_2| \leq K_2 \gamma$.

Step 3 (B). Let $\tilde{F} = T F$ ($|\tilde{F}| \leq 1$) and $H = (I - T_0)f_0$ ($|H| \leq 2$); then $|E_2 - E_3| = |\mathbb{E}[\tilde{F}(A)H(B)T f_m(C)]|$. (i) $\|\tilde{F}^{> d_1}\|_2 \leq (1 - \gamma)^{d_1}$ (an ES component of degree e involves at least e bits), so replacing $\tilde{F}$ by $\tilde{F}^{\leq d_1}$ costs at most $2(1 - \gamma)^{d_1}$. (ii) $\|H^{\leq d_2}\|_2 \leq d_2 \gamma_0$, so by Cauchy-Schwarz $|\mathbb{E}[\tilde{F}^{\leq d_1} H^{\leq d_2} T f_m]| \leq \|\tilde{F}^{\leq d_1}\|_2 \|H^{\leq d_2}\|_2 \leq d_2 \gamma_0$. (iii) Let $\Psi = \tilde{F}^{\leq d_1}(A)H^{> d_2}(B)$, a function on the (A, B) -space with no ES components of degree $\leq d_2 - d_1$, and let $\mathcal{T}_3$ map functions of (A, B) to functions of C . Then $|\mathbb{E}[\Psi T f_m(C)]| \leq \|\mathcal{T}_3 \Psi\|_2 \leq r_2^{d_2 - d_1 + 1} \|\Psi\|_2$, and

$$\|\Psi\|_2 \leq \|\tilde{F}^{\leq d_1} H\|_2 + \|\tilde{F}^{\leq d_1} H^{\leq d_2}\|_2 \leq 2 + \|\tilde{F}^{\leq d_1}\|_{2q} \|H^{\leq d_2}\|_{2q/(q-1)} \leq 2 + 2(2qM_A)^{d_1/2} \beta_q^{d_2},$$

by Hölder ($\frac{1}{2q} + \frac{q-1}{2q} = \frac{1}{2}$) and Lemma 6.6, where $\beta_q = \sqrt{\frac{q+1}{q-1}} M_B^{1/(2q)}$. Fix q so large that $r_2 \beta_q \leq \sqrt{r_2}$.

Then (iii) is at most $4(\sqrt{2qM_A}/r_2)^{d_1} r_2^{d_2/2}$. Altogether

$$|E_0 - E_3| \leq r_1^{D+1} + (m-1)D\gamma + K_2\gamma + 2(1 - \gamma)^{d_1} + d_2\gamma_0 + 4(\sqrt{2qM_A}/r_2)^{d_1} r_2^{d_2/2}.$$

Choose, in this order: D with $r_1^{D+1} \leq \epsilon/8$; γ with $(m-1)D\gamma \leq \epsilon/8$ and $K_2\gamma \leq \epsilon/4$; d_1 with $2(1 - \gamma)^{d_1} \leq \epsilon/8$; d_2 with $4(\sqrt{2qM_A}/r_2)^{d_1} r_2^{d_2/2} \leq \epsilon/8$; γ_0 with $d_2\gamma_0 \leq \epsilon/8$. Then $|E_0 - E_3| \leq \frac{7}{8}\epsilon$. None of the choices depends on n or on the functions. $\square$

6.3 Proof of Theorem 6.1

Let $\epsilon > 0$ and choose $\epsilon_1, \dots, \epsilon_4 > 0$ with $2(m+1)\epsilon_4 + \|\hat{g}\|_1(\epsilon_1 + (2m+1)\epsilon_2 + 2m\epsilon_3) \leq 2\epsilon$. Let γ, γ_0 be given by Lemma 6.7 for ϵ_1 ; let (d, τ) be given by Theorem 6.3 for the distribution $\mathcal{D}$ of $(x_1, \dots, x_m, g(x))$, accuracy ϵ_2 and noise bound $\min(\gamma, \gamma_0)$; let L be given by Theorem 6.2 for (d, τ, ϵ_4) ; and let $\delta = 2^{-L}\epsilon_3$. By Remark 1.2(b) it suffices to show that every (δ, L) -pseudorandom $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$ satisfies $\Pr[f \circ g^n = g \circ f^m] \leq s_g^{\text{lab}} + \epsilon$.

Let Λ be given by Theorem 6.2, $k = |\Lambda| \leq L$, and for $z \in \{-1, 1\}^\Lambda$ let $f_z = f_{\Lambda \rightarrow z}$ (a function on $\{-1, 1\}^{[n] \setminus \Lambda}$). Call z *good* if f_z is (d, τ) -regular w.r.t. π_0 , and μ -*good* if it is regular w.r.t. π_μ . Let $W = Z|_{\Lambda \times [m]}$; it is uniform, its columns are uniform and $g(W) \sim \pi_\mu^\Lambda$, so W is *good* (all columns good and $g(W)$ μ -good) with probability $\geq 1 - (m+1)\epsilon_4$. For every z , $|\mathbb{E}f_z| = |\sum_{S \subseteq \Lambda} \hat{f}(S)\chi_S(z)| \leq 2^L\delta = \epsilon_3$.

For $z \in \{-1, 1\}^\Lambda$ define on $\mathbb{R}^{[n] \setminus \Lambda}$: $\tilde{Q}_z = \text{clip}(T_{1-\gamma}f_z)$ (multilinear extension), and $Q_z = (\tilde{Q}_z - \mathbb{E}\tilde{Q}_z)/(1 + |\mathbb{E}\tilde{Q}_z|)$ if z is good, $Q_z \equiv 0$ otherwise; and $\tilde{P}_\nu(x) = \text{clip}(T_{1-\gamma_0}^\mu f_\nu)(\mu + \sqrt{1-\mu^2}x)$. Then $(Q_z)_z$ is an admissible family in Definition 1.5 (with $n' = n - k$), and for good z , $|\mathbb{E}\tilde{Q}_z| \leq |\mathbb{E}f_z| + \epsilon_2 \leq \epsilon_2 + \epsilon_3$ by Theorem 6.3.

Fix a good W and $S \subseteq [m]$. Conditioned on W , the remaining rows of Z are i.i.d. uniform, so with $(x_1, \dots, x_m, y) \sim \mathcal{D}^{[n] \setminus \Lambda}$,

$$\mathbb{E}_{Z|W} \left[f(g^n(Z)) \prod_{j \in S} f(\text{col}_j Z) \right] = \mathbb{E} \left[f_{g(W)}(y) \prod_{j \in S} f_{W_j}(x_j) \right].$$

Apply Lemma 6.7 with $F = \prod_{j \in S, j < m} f_{W_j}$, $f_m = f_{W_m}$ if $m \in S$ and $f_m \equiv 1$ otherwise, $f_0 = f_{g(W)}$ (error $\leq \epsilon_1$); then Theorem 6.3 to the regular functions f_{W_j} ($j \in S$, rate γ), $f_{g(W)}$ (rate γ_0) and the constant 1 for $j \notin S$ (error $< \epsilon_2$); the Gaussian analogue of $\mathcal{D}$ is $(\mathcal{G}_1, \dots, \mathcal{G}_m, \mu + \sqrt{1-\mu^2}\mathcal{G}_0)$ with $(\mathcal{G}_0, \dots, \mathcal{G}_m) \sim \mathcal{N}_g$ (CFMMS Lemma 5.10). This gives

$$\left| \mathbb{E}_{Z|W} \left[f(g^n Z) \prod_{j \in S} f(\text{col}_j Z) \right] - \mathbb{E} \left[\tilde{P}_{g(W)}(\mathcal{G}_0) \prod_{j \in S} \tilde{Q}_{W_j}(\mathcal{G}_j) \right] \right| \leq \epsilon_1 + \epsilon_2.$$

Replacing $\tilde{Q}_{W_j}$ by $\tilde{Q}_{W_j}/(1+|e_j|)$ and then by Q_{W_j} ($e_j = \mathbb{E}\tilde{Q}_{W_j}$) changes the Gaussian expectation by at most $\sum_j |e_j| + \sum_j \frac{|e_j|}{1+|e_j|} \leq 2m(\epsilon_2 + \epsilon_3)$ (all factors are bounded by 1). Multiplying by $\hat{g}(S)$ and summing over S :

$$\left| \mathbb{E}_{Z|W} [(f \circ g^n)(g \circ f^m)] - \mathbb{E}_{\mathcal{G}} \left[\tilde{P}_{g(W)}(\mathcal{G}_0) g(Q_{W_1}(\mathcal{G}_1), \dots, Q_{W_m}(\mathcal{G}_m)) \right] \right| \leq \|\hat{g}\|_1(\epsilon_1 + (2m+1)\epsilon_2 + 2m\epsilon_3)$$

for good W ; for bad W the difference is at most 2. Averaging over W (which is independent of $\mathcal{G}$),

$$\begin{aligned} \mathbb{E}_Z [(f \circ g^n)(g \circ f^m)] &\leq \sum_{\nu} \Pr[g(W) = \nu] \mathbb{E}_x [\tilde{P}_\nu(x) \gamma_\nu(x)] + 2\epsilon \\ &\leq \sum_{\nu} \Pr[g(W) = \nu] \mathbb{E}_x |\gamma_\nu(x)| + 2\epsilon \leq 2s_g^{\text{lab}} - 1 + 2\epsilon, \end{aligned}$$

where γ_ν is computed for the family (Q_z) . Hence $\Pr[f \circ g^n = g \circ f^m] \leq s_g^{\text{lab}} + \epsilon$. $\square$

Theorems 5.1 and 6.1 together with Lemma 2.3(c) prove Theorem B.

7 Nontriviality: proof of Theorem C

Throughout, g is as in Theorem B. Note $\rho_j^2 < 1$ for all j ($\rho_j^2 = 1$ would force g to depend on x_j only).

(i) By Lemma 2.3(b), for mean-zero $[-1, 1]$ -valued q_j , $\gamma_q(x) = \mathbb{E}[h(Y) \mid \mathcal{G}_0 = x]$ with $Y \in [-1, 1]^J$; if $\text{sgn}(\mu)h \geq 0$ on $\{-1, 1\}^J$ then also on $[-1, 1]^J$ by (M1), so $\text{sgn}(\mu)\gamma_q \geq 0$ and $\text{val}(q) = \text{triv}_g$ by Lemma 2.3(a). Thus $s_g^U = \text{triv}_g$, and $\text{triv}_g \leq s_g^L \leq s_g \leq s_g^U$ gives (i).

(ii) Let $\text{sgn}(\mu)h(w) < 0$. Take $n = 1$, $q_j(y) = w_j \text{sgn}(\rho_j) \text{sgn}(y)$ for $j \in J$ and $q_j = \text{sgn}$ for $j \in U$ (all ± 1 -valued, mean zero). Given $\mathcal{G}_0 = x$, $\Pr[\text{sgn} \mathcal{G}_j = \text{sgn}(\rho_j x)] = \Phi(|\rho_j x| / \sqrt{1 - \rho_j^2}) \rightarrow 1$ as $x \rightarrow +\infty$, so by Lemma 2.3(b) and (M2), $\gamma_q(x) \rightarrow h(w)$. Hence $\text{sgn}(\mu)\gamma_q < 0$ on some (x_1, ∞) and $\text{val}(q) > \text{triv}_g$ by Lemma 2.3(a). With (i) this gives the equivalence.

(iii) Let $\text{sgn}(\mu)\Gamma_\nu(F) = -\theta < 0$. Let ρ_c denote the common value of ρ_j on class c (all $\rho_c \neq 0$, pairwise distinct). Take $n = 1$; for $R, X > 0$ let $I_c = [\rho_c X - R - 1, \rho_c X + R + 1]$. Choose R with $8d\Phi(-R) \leq \theta$ and then X so large that the intervals I_c are pairwise disjoint and $\sum_c \gamma_1(I_c) \leq \frac{1}{2}$. Define

$$Q_z(y) = F_c(z) \quad (y \in I_c), \quad Q_z(y) = \kappa_z := -\frac{\sum_c F_c(z)\gamma_1(I_c)}{1 - \sum_c \gamma_1(I_c)} \quad (y \notin \bigcup_c I_c).$$

Then $|Q_z| \leq 1$ and $\mathbb{E}Q_z = 0$. For $x \in [X, X + 1]$ and $j \in J$, $\mathcal{G}_j \mid x \sim N(\rho_j x, 1 - \rho_j^2)$ with $|\rho_j x - \rho_j X| \leq 1$, so $\Pr[\mathcal{G}_j \notin I_{c(j)} \mid x] \leq 2\Phi(-R)$. On the event that $\mathcal{G}_j \in I_{c(j)}$ for all $j \in J$ (independent of the labels) the J -inputs equal $F_{c(j)}(C_j)$; by Lemma 2.3(b) and $|h| \leq 1$, $|\gamma_\nu(x) - \Gamma_\nu(F)| \leq 2 \cdot 2d\Phi(-R) \leq \theta/2$. So $\text{sgn}(\mu)\gamma_\nu \leq -\theta/2$ on $[X, X + 1]$, and $\text{val}^{\text{lab}}(Q) > \text{triv}_g$ by Lemma 2.3(a) ($\Pr[g(W) = \nu] > 0$ as g is nonconstant). By Theorem 5.1, $s_g > \text{triv}_g$. For $k = 0$ the family is a single function, so $s_g^L \geq \text{val}(Q, \dots, Q) > \text{triv}_g$.

(iv) Take $k = 1$ and $F_c = \text{id}$ for all classes, and $\nu = -\text{sgn}(\mu)$. Since $\mathbb{E}_{\pi_b} h = \frac{\mathbb{E}[h + bh^2]}{1 + b\mu} = \frac{\mu + b\|h\|_2^2}{1 + b\mu}$, we get $\text{sgn}(\mu)\Gamma_\nu(F) = \frac{|\mu| - \|h\|_2^2}{1 - |\mu|} < 0$; apply (iii). (Explicitly: $n = 1$, $Q_z(y) = z \text{sgn}(|y| - \Phi^{-1}(3/4))$ works, without using classes.) If all level-one coefficients are nonzero then $J = [m]$, $h = g$ and $\|h\|^2 = 1 > |\mu|$.

(v) See §7.1.

7.1 Small arity: proof of Theorem C(v)

The script `numerics/q6_exact.py` enumerates all $g: \{-1, 1\}^m \rightarrow \{-1, 1\}$, $m \leq 4$, with $\mathbb{E}g \neq 0$, depending on all coordinates and $g \neq \pm \prod x_j$ (no symmetry reduction), computes J , h and the classes exactly (integers), and certifies each g by one of: TRIV ($\text{sgn}(\mu)h \geq 0$; then everything is trivial by (i)); L0 (a class-constant $w \in \{-1, 1\}^J$ with $\text{sgn}(\mu)h(w) < 0$; then s_g^L , hence s_g , is nontrivial by (iii), $k = 0$); LAB1 ($k = 1$, $\nu \in \{\pm 1\}$ and class-wise decoders $F_c \in \{\text{id}, -\text{id}, +1, -1\}$ with $\text{sgn}(\mu)\Gamma_\nu(F) < 0$; then s_g is nontrivial by (iii)). Result:

m	functions	TRIV	L0	LAB1	undecided
2	8	0	8	0	0
3	160	32	124	4	0
4	51976	5448	44244	2284	0

The undecided count is 0: every g that is not TRIV, i.e. every g with $\text{sgn}(\mu)h(w) < 0$ for some w (equivalently, by (ii), $s_g^U > \text{triv}_g$), received an L0 or LAB1 certificate and so has $s_g > \text{triv}_g$, while every TRIV g has $s_g = s_g^U = \text{triv}_g$ by (i). This proves (v). At $m = 3$ the four LAB1 functions are $\pm \text{E2}(\pm x)$, the orbit of E2.

8 Labels absorb the Gaussian part: proof of Theorem D

(a) $\iff$ (b) is Theorem B. (c) $\Rightarrow$ (b) is contained in the proof of Theorem C(iii), which constructs a family (Q_z) with $\text{val}^{\text{lab}}(Q) > \text{triv}_g$, and (d) $\Rightarrow$ (c) follows by taking $F_c = F$ for every class; so the equivalence of (b), (c) and (d) does not use §5 and §6. For the rationality claim in (d): π_b has rational values, so $\Gamma_\nu(F)$ is a polynomial with rational coefficients in the values $F(z)$, $z \in \{-1, 1\}^k$, and a strict inequality survives a small perturbation of F to rational values in $[-1, 1]$. It remains to prove (b) $\Rightarrow$ (d). The idea is that the Gaussian part of a labelled configuration is itself a limit of normalised sums of further label rows (multivariate CLT), so every sign change of some γ_ν at a finite point is an average of labelled values “at infinity”.

Step 0. $\sum_j \rho_j^2 < 1$. Indeed $\sum_j \rho_j^2 = \sum_j \hat{g}(\{j\})^2 / (1 - \mu^2)$ and $\sum_{S \neq \emptyset} \hat{g}(S)^2 = 1 - \mu^2$, so equality would mean that $g = \mu + \sum_j \hat{g}(\{j\})x_j$ has degree ≤ 1 . A $\{-1, 1\}$ -valued function of degree ≤ 1 is constant or $\pm x_j$ (if $\hat{g}(\{i\}), \hat{g}(\{j\}) \neq 0$ for $i \neq j$, the four values $c \pm \hat{g}(\{i\}) \pm \hat{g}(\{j\})$ obtained by flipping x_i, x_j lie in $\{-1, 1\}$, which forces $|\hat{g}(\{i\})| = |\hat{g}(\{j\})| = 1$, contradicting Parseval), and both are excluded by $\mu \neq 0$ and g nonconstant.

Step 1 (a box where γ_ν has the wrong sign). Assume (b). By Lemma 2.3(a) there are k, n , a family (Q_z) and $\nu_0 \in \{-1, 1\}^k$ such that $\sigma\gamma_{\nu_0} < 0$ on a set of positive measure. Then $d \geq 1$: if $J = \emptyset$, then $h \equiv \mu$ and Lemma 2.3(b) gives $\gamma_{\nu_0} \equiv \mu$. By Lemma 2.1 and Step 0, conditionally on $\mathcal{G}_0 = x$ the vector $(\mathcal{G}_j)_{j \in J}$ is Gaussian with mean $(\rho_j x)_{j \in J}$ and nonsingular covariance $(I_J - \rho_J \rho_J^\top) \otimes I_n$; let ψ be the density of the centred Gaussian with this covariance on $(\mathbb{R}^n)^J$. By Lemma 2.3(b),

$$\gamma_{\nu_0}(x) = \sum_M \Pr[C = M] \int h((Q_{M,j}(y_j))_{j \in J}) \psi(y - (\rho_j x)_{j \in J}) dy,$$

which is continuous in x because translation is continuous in L^1 and $|h| \leq 1$. Hence there are a closed box $B = x_0 + [-\varepsilon, \varepsilon]^n$ ($\varepsilon > 0$) and $\theta > 0$ with $\sigma\gamma_{\nu_0} \leq -\theta$ on B , so

$$\sigma \mathbb{E}[\gamma_{\nu_0}(\mathcal{G}_0) \mathbf{1}\{\mathcal{G}_0 \in B\}] \leq -\theta \gamma_n(B) < 0. \quad (7)$$

Step 2 (continuous decoders). Let $\eta = \theta\gamma_n(B)/(3d)$ and choose continuous $\tilde{Q}_z: \mathbb{R}^n \rightarrow [-1, 1]$ with $\|Q_z - \tilde{Q}_z\|_{L^1(\gamma_n)} \leq \eta$ (continuous functions are dense in $L^1(\gamma_n)$, and truncation to $[-1, 1]$ does not increase the distance to a $[-1, 1]$ -valued function; we do not need $\mathbb{E}\tilde{Q}_z = 0$). For $M \in \{-1, 1\}^{k \times J}$ define the bounded function

$$\phi_M(s_0, (s_j)_{j \in J}) = h((\tilde{Q}_{M,j}(s_j))_{j \in J}) \mathbf{1}\{s_0 \in B\} \quad (s_0, s_j \in \mathbb{R}^n).$$

Let C be as in Lemma 2.3(b) for ν_0 , independent of the $\mathcal{G}$'s. By Lemma 2.3(b) and the tower property, $\mathbb{E}[\gamma_{\nu_0}(\mathcal{G}_0) \mathbf{1}\{\mathcal{G}_0 \in B\}] = \mathbb{E}[h((Q_{C,j}(\mathcal{G}_j))_j) \mathbf{1}\{\mathcal{G}_0 \in B\}]$, and by (M2), since each $\mathcal{G}_j \sim N(0, I_n)$ is independent of C , this differs from $\mathbb{E}[\phi_C(\mathcal{G}_0, \mathcal{G}_J)]$ by at most $\sum_{j \in J} \mathbb{E}|Q_{C,j}(\mathcal{G}_j) - \tilde{Q}_{C,j}(\mathcal{G}_j)| \leq d\eta$. With (7),

$$\sigma \mathbb{E}[\phi_C(\mathcal{G}_0, \mathcal{G}_J)] \leq -\frac{2}{3} \theta \gamma_n(B). \quad (8)$$

Step 3 (CLT). Let $N \in \mathbb{N}$ and let W' be uniform in $\{-1, 1\}^{(nN) \times m}$, with rows indexed by $(r, t) \in [n] \times [N]$, independent of everything else; put $\nu' = g(W') \in \{-1, 1\}^{nN}$ (row-wise) and, for $z' \in \mathbb{R}^{[n] \times [N]}$, $\Sigma_N(z') \in \mathbb{R}^n$, $\Sigma_N(z')_r = N^{-1/2} \sum_{t=1}^N z'_{(r,t)}$. Let

$$S_j = \Sigma_N(W'_{\cdot j}) \quad (j \in J), \quad S_0 = \Sigma_N((\nu' - \mu)/\sqrt{1 - \mu^2}).$$

The vectors $\xi_{(r,t)} = ((g(W'_{(r,t),\cdot}) - \mu)/\sqrt{1-\mu^2}, (W'_{(r,t),j})_{j \in J}) \in \mathbb{R}^{1+d}$ are i.i.d. and bounded; writing $\xi = (\xi_0, (\xi_j)_{j \in J})$ for any one of them, they have mean 0 and covariance $\mathbb{E}[\xi_0^2] = \mathbb{E}[\xi_j^2] = 1$, $\mathbb{E}[\xi_j \xi_l] = 0$ ($j \neq l$), $\mathbb{E}[\xi_0 \xi_j] = \hat{g}(\{j\})/\sqrt{1-\mu^2} = \rho_j$: the covariance of $(\mathcal{G}_0, (\mathcal{G}_j)_{j \in J})$ under $\mathcal{N}_g$. By the multivariate central limit theorem, applied to each $r \in [n]$ (different r are independent), (S_0, S_J) converges in distribution to $(\mathcal{G}_0, \mathcal{G}_J)$ as $N \rightarrow \infty$. For each M the function ϕ_M is bounded and continuous off $\partial B \times (\mathbb{R}^n)^J$, a null set for the law of $(\mathcal{G}_0, \mathcal{G}_J)$; by the portmanteau theorem $\mathbb{E}\phi_M(S_0, S_J) \rightarrow \mathbb{E}\phi_M(\mathcal{G}_0, \mathcal{G}_J)$. Taking C independent of W' and averaging over M ,

$$\mathbb{E}[\phi_C(S_0, S_J)] \longrightarrow \mathbb{E}[\phi_C(\mathcal{G}_0, \mathcal{G}_J)] \quad (N \rightarrow \infty). \quad (9)$$

Step 4 (identification). Let W be uniform in $\{-1, 1\}^{k \times m}$, independent of W' , and consider the matrix $\bar{W} = \begin{pmatrix} W \\ W' \end{pmatrix}$ with $k + nN$ rows, which is uniform; its output is $g(\bar{W}) = (g(W), \nu')$. Define the single decoder

$$F_N: \{-1, 1\}^k \times \{-1, 1\}^{[n] \times [N]} \rightarrow [-1, 1], \quad F_N(z, z') = \tilde{Q}_z(\Sigma_N(z')),$$

so that $F_N(\bar{W}_{\cdot j}) = \tilde{Q}_{W_{\cdot j}}(S_j)$. Conditionally on $g(W) = \nu_0$, $W_{\cdot j}$ has the law of C (Lemma 2.3(b)) and is independent of W' , so $\mathbb{E}[\phi_C(S_0, S_J)] = \mathbb{E}[\phi_{W_{\cdot j}}(S_0, S_J) \mid g(W) = \nu_0]$. Since S_0 is a function $s_0(\nu')$ of ν' and ν' is independent of W ,

$$\mathbb{E}[\phi_C(S_0, S_J)] = \sum_{\nu': s_0(\nu') \in B} \Pr[g(W') = \nu'] \mathbb{E}\left[h((F_N(\bar{W}_{\cdot j}))_{j \in J}) \mid g(\bar{W}) = (\nu_0, \nu')\right].$$

As in the proof of Lemma 2.3(b) (rows of $\bar{W}$ are independent and the output of a row depends on that row only), conditionally on $g(\bar{W}) = \bar{\nu}$ the J -columns of $\bar{W}$ have independent rows with laws $\pi_{\bar{\nu}_i}$; hence the conditional expectation is exactly the quantity in (d) for $k + nN$ label bits, output $\bar{\nu} = (\nu_0, \nu')$ and the single decoder F_N . By (8) and (9), $\sigma \mathbb{E}[\phi_C(S_0, S_J)] < 0$ for N large, so at least one term of the (finite, nonnegatively weighted) sum has the sign required in (d). $\square$

Remark 8.1. Theorem D makes the finite- x structure of the labelled problem (the negative correlations between the inputs given $\mathcal{G}_0 = x$) irrelevant for nontriviality: a sign change anywhere is already visible at $x = \infty$ with more labels. It also shows that the class-wise decoders of Theorem C(iii) are never needed; they remain convenient for writing down short certificates. This can also be seen directly: append K label rows with output $+1$. The extra entries of column j are i.i.d. with mean $\hat{h}(\{j\})/(1 + \mu)$, which differs between classes, so the single decoder that applies F_c to the first k entries when the empirical mean of the other K is closest to the mean of class c gives a value that differs from $\Gamma_\nu(F)$ by at most twice the probability of misclassifying some column, which tends to 0 as $K \rightarrow \infty$. By contrast, Theorem D says nothing about the *value* of s_g : dictator decoders give large “values at infinity” that are not realisable.

9 An explicit g with $s_g = \text{triv}_g < s_g^U$: proof of Theorem E

9.1 The function g_0

Write the inputs as $(w_1, w_2, w_3, u_1, u_2)$ and let

$$\begin{aligned} g_0(w, u) = +1 &\iff (u_1, u_2) \in A_w, \quad \text{where} \\ A_{+++} = A_{++-} = A_{+-+} &= \emptyset, \quad A_{+--} = \{\pm 1\}^2 \setminus \{(+, +)\}, \\ A_{-++} = \{(+, +), (-, -)\}, \quad A_{-+-} &= \{(+, +)\}, \quad A_{--+} = \{(+, -)\}, \quad A_{---} = \{(-, +)\}. \end{aligned} \quad (10)$$

Then $h(w) = \mathbb{E}_u g_0(w, u) = \frac{1}{2}|A_w| - 1$, that is, in the order $+++$, $++-$, $+ - +$, $+ - -$, $- + +$, $- + -$, $- - +$, $- - -$,

$$h = (-1, -1, -1, \frac{1}{2}, 0, -\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}), \quad \mu = \mathbb{E}h = -\frac{1}{2}, \quad \sigma = -1, \quad \text{triv}_{g_0} = \frac{3}{4}.$$

A direct count gives $\hat{g}_0(\{w_j\}) = \hat{h}(\{j\}) = -\frac{1}{8}$ for $j = 1, 2, 3$ and

$$\hat{g}_0(\{u_i\}) = 2 \cdot 2^{-5} \sum_w \sum_{(u_1, u_2) \in A_w} u_i = 0 \quad (i = 1, 2)$$

(over $w = + - -, - + +, - + -, - - +, - - -$ the inner sums are $-1, 0, 1, 1, -1$ for $i = 1$ and $-1, 0, 1, -1, 1$ for $i = 2$). So $J = \{1, 2, 3\}$, and J is a single class. The function g_0 depends on u_1 and u_2 (because of A_{-++}) and on each w_j (because $\hat{g}_0(\{w_j\}) \neq 0$), it is not $\pm$ a parity and $\mu \neq 0$, so Theorems B–D apply. The two row laws are

$$\pi_+ = \frac{1+h}{4} = (0, 0, 0, \frac{3}{8}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}), \quad \pi_- = \frac{1-h}{12} = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{24}, \frac{1}{12}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}).$$

For $k \geq 0$, $\nu \in \{-1, 1\}^k$ and $F: \{-1, 1\}^k \rightarrow [-1, 1]$ let $C \in \{-1, 1\}^{k \times 3}$ have independent rows, row $i \sim \pi_{\nu_i}$, and let $X, Y, Z \in \{-1, 1\}^k$ be its three columns, so that $\Gamma_\nu(F) = \mathbb{E} h(F(X), F(Y), F(Z))$.

Proposition 9.1. $\Gamma_\nu(F) \leq 0$ for all $k \geq 0$, $\nu \in \{-1, 1\}^k$ and $F: \{-1, 1\}^k \rightarrow [-1, 1]$.

Proof of Theorem E and Corollary E', given Proposition 9.1. Since $\sigma = -1$, Proposition 9.1 says that condition (d) of Theorem D fails for g_0 , so $s_{g_0} \leq \text{triv}_{g_0}$ by Theorem D. With Theorem B and Lemma 2.3(a), $\text{triv}_{g_0} \leq s_{g_0}^L \leq s_{g_0} \leq \text{triv}_{g_0}$. Finally $\sigma h(+, -, -) = -\frac{1}{2} < 0$, so $s_{g_0}^U > \text{triv}_{g_0}$ by Theorem C(ii).

For $m \geq 5$ let $g_m(w, u_1, \dots, u_{m-3}) = g_0(w, u_1, u_2 u_3 \cdots u_{m-3})$ (so $g_5 = g_0$). The product $u' = u_2 \cdots u_{m-3}$ is uniform and independent of u_1 , so g_m has the same h , the same μ and the same level-one coefficients $\hat{g}_m(\{w_j\}) = -\frac{1}{8}$; for $i \geq 2$, u_i is uniform and independent of (w, u_1, u') when $m \geq 6$, so $\hat{g}_m(\{u_i\}) = \mathbb{E}[g_0(w, u_1, u')]\mathbb{E}[u_i] = 0$, and $\hat{g}_m(\{u_1\}) = \hat{g}_0(\{u_1\}) = 0$. As g_0 depends on u_2 , g_m depends on every u_i . Thus g_m satisfies the hypotheses of Theorems B–D with the same J , h and σ as g_0 . Conditions (d) of Theorem D and the condition of Theorem C(ii) depend only on h and σ , so the argument above applies verbatim. For $m \leq 4$ and $\mathbb{E}g \neq 0$, Theorem C(v) excludes $s_g = \text{triv}_g < s_g^U$. $\square$

9.2 Reduction to a quadratic inequality

Lemma 9.2. Let $P = \frac{1+F}{2}: \{-1, 1\}^k \rightarrow [0, 1]$ and $P_1 = P(X)$, $P_2 = P(Y)$, $P_3 = P(Z)$. Then

$$2\Gamma_\nu(F) \leq 2\mathbb{E}P_1 - 4\mathbb{E}[P_1P_2] + \mathbb{E}[P_2P_3] - 1.$$

Proof. For $y \in [-1, 1]^3$ put $p_j = \frac{1+y_j}{2}$. Then

$$2h(y) = -1 + 2p_1 - 3p_1p_2 - 3p_1p_3 + p_2p_3 + 2p_1p_2p_3,$$

since both sides are multilinear in p and they agree at the eight vertices $p \in \{0, 1\}^3$ (the right-hand side takes the values $-2, -2, -2, 1, 0, -1, -1, -1$ at $p = (1, 1, 1), (1, 1, 0), \dots, (0, 0, 0)$). Given C the three inputs are $F(X), F(Y), F(Z)$, so $2\Gamma_\nu(F) = \mathbb{E}[-1 + 2P_1 - 3P_1P_2 - 3P_1P_3 + P_2P_3 + 2P_1P_2P_3]$. Both π_+ and π_- are invariant under exchanging w_2 and w_3 (their values at $++-$ and $+ - +$ agree, as do their values at $- + -$ and $- - +$), and the rows of C are independent, so (X, Y, Z) and (X, Z, Y) have the same law and $\mathbb{E}[P_1P_3] = \mathbb{E}[P_1P_2]$. Finally $P_1P_2P_3 \leq P_1P_2$ since $0 \leq P_3 \leq 1$ and $P_1P_2 \geq 0$. $\square$

9.3 Spectral form

Under π_b each coordinate w_j equals $+1$ with probability β_b , where $\beta_+ = \frac{3}{8}$ and $\beta_- = \frac{13}{24}$ (in general $\mathbb{E}_{\pi_b} w_j = b\hat{h}(\{j\})/(1 + b\mu)$, which here does not depend on j). Hence X, Y, Z all have the law of a random point x of $\Omega_\nu = \{-1, 1\}^k$ with independent coordinates and $\Pr[x_i = 1] = \beta_{\nu_i}$. For $\beta \in (0, 1)$ let $\chi^{(\beta)}(t) = (t - (2\beta - 1))/(2\sqrt{\beta(1 - \beta)})$ for $t \in \{-1, 1\}$, which has mean 0 and variance 1 when $\Pr[t = 1] = \beta$. Put $\chi_i(x) = \chi^{(\beta_{\nu_i})}(x_i)$ and $\chi_S = \prod_{i \in S} \chi_i$; these form an orthonormal basis of $L^2(\Omega_\nu)$, and we write $P = \sum_S \hat{P}(S) \chi_S$. Then $\hat{P}(\emptyset) = \mathbb{E}P =: p$ and $\sum_{S \neq \emptyset} \hat{P}(S)^2 = \text{Var } P \leq p - p^2$ (as $P^2 \leq P$). Let

$$W_+ = \sum_{i: \nu_i = +1} \hat{P}(\{i\})^2, \quad a^* = \left(\frac{7}{15}\right)^3 + 4\left(\frac{3}{5}\right)^3 = \frac{3259}{3375}.$$

Lemma 9.3. $2\mathbb{E}P_1 - 4\mathbb{E}[P_1 P_2] + \mathbb{E}[P_2 P_3] \leq 2p - 3p^2 + a^*(p - p^2) + (\frac{43}{15} - a^*)W_+.$

Proof. The law of (w_1, w_2) on $(+, +, +, -, -, -)$ is $(0, \frac{3}{8}, \frac{3}{8}, \frac{1}{4})$ under π_+ and $(\frac{1}{3}, \frac{5}{24}, \frac{5}{24}, \frac{1}{4})$ under π_- ; the law of (w_2, w_3) is $(\frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{2})$ under π_+ and $(\frac{1}{4}, \frac{7}{24}, \frac{7}{24}, \frac{1}{6})$ under π_- . For $t, t' \in \{-1, 1\}$ with $\Pr[t = 1] = \Pr[t' = 1] = \beta$ we have $\mathbb{E}[\chi^{(\beta)}(t)\chi^{(\beta)}(t')] = (\mathbb{E}[tt'] - (2\beta - 1)^2)/(1 - (2\beta - 1)^2)$, so the correlations $\lambda_b^{12} = \mathbb{E}_{\pi_b}[\chi^{(\beta_b)}(w_1)\chi^{(\beta_b)}(w_2)]$ and λ_b^{23} (defined likewise with w_2, w_3) are

$$\lambda_+^{12} = \frac{-\frac{1}{2} - \frac{1}{16}}{\frac{15}{16}} = -\frac{3}{5}, \quad \lambda_-^{12} = \frac{\frac{1}{6} - \frac{1}{144}}{\frac{143}{144}} = \frac{23}{143}, \quad \lambda_+^{23} = \frac{\frac{1}{2} - \frac{1}{16}}{\frac{15}{16}} = \frac{7}{15}, \quad \lambda_-^{23} = \frac{-\frac{1}{6} - \frac{1}{144}}{\frac{143}{144}} = -\frac{25}{143}.$$

Let (x, x') be a pair of random points of $\{-1, 1\}^k$ such that the pairs (x_i, x'_i) are independent over i , both x and x' have the law of Ω_ν , and $\mathbb{E}[\chi_i(x)\chi_i(x')] = \lambda_i$. For $S \neq T$ pick $i \in S \Delta T$; the factor of $\mathbb{E}[\chi_S(x)\chi_T(x')]$ belonging to coordinate i is $\mathbb{E}\chi_i(x) = 0$ or $\mathbb{E}\chi_i(x') = 0$. Hence $\mathbb{E}[\chi_S(x)\chi_T(x')] = \mathbf{1}\{S = T\} \prod_{i \in S} \lambda_i$ and $\mathbb{E}[P(x)P(x')] = \sum_S \lambda_S \hat{P}(S)^2$ with $\lambda_S = \prod_{i \in S} \lambda_i$. Applying this to (X, Y) and to (Y, Z) , with $\lambda_S^{jl} = \prod_{i \in S} \lambda_{\nu_i}^{jl}$,

$$2\mathbb{E}P_1 - 4\mathbb{E}[P_1 P_2] + \mathbb{E}[P_2 P_3] = 2p - 3p^2 + \sum_{S \neq \emptyset} a_S \hat{P}(S)^2, \quad a_S = \lambda_S^{23} - 4\lambda_S^{12}.$$

If $S = \{i\}$ with $\nu_i = +1$ then $a_S = \frac{7}{15} + \frac{12}{5} = \frac{43}{15}$. We claim $a_S \leq a^*$ for every other nonempty S . If $S = \{i\}$ with $\nu_i = -1$, $a_S = -\frac{25}{143} - \frac{92}{143} < 0$. If $|S| = 2$, a_S is $\frac{49}{225} - \frac{36}{25} = -\frac{11}{9}$, or $-\frac{7}{15} \cdot \frac{25}{143} + 4 \cdot \frac{3}{5} \cdot \frac{23}{143} = \frac{653}{2145}$, or $\frac{625 - 2116}{20449} < 0$, according as S contains two, one or no indices i with $\nu_i = +1$; all are below a^* . If $|S| \geq 3$, then, as $|\lambda_b^{23}| \leq \frac{7}{15}$ and $|\lambda_b^{12}| \leq \frac{3}{5}$ for both b , $a_S \leq |\lambda_S^{23}| + 4|\lambda_S^{12}| \leq (\frac{7}{15})^{|S|} + 4(\frac{3}{5})^{|S|} \leq a^*$. Therefore, using $a^* \geq 0$,

$$\sum_{S \neq \emptyset} a_S \hat{P}(S)^2 \leq \frac{43}{15} W_+ + a^*(\text{Var } P - W_+) \leq a^*(p - p^2) + (\frac{43}{15} - a^*)W_+. \quad \square$$

The only coefficient exceeding 1 is the one on the level-one weight of the $\nu = +1$ rows, and W_+ is the level-one weight of a $[0, 1]$ -valued function of $\frac{3}{8}$ -biased bits: for i with $\nu_i = +1$, $\hat{P}(\{i\}) = \mathbb{E}[P(x)\chi_i(x)] = \mathbb{E}[\bar{P}(x_+)\chi_i(x)]$, where $x_+ = (x_i)_{\nu_i = +1}$ and $\bar{P}(x_+) = \mathbb{E}[P(x) \mid x_+] \in [0, 1]$. So $W_+ = W^1(\bar{P})$, the level-one weight of $\bar{P}: \{-1, 1\}^{k_+} \rightarrow [0, 1]$ on the $\frac{3}{8}$ -biased cube. Theorem K bounds it.

9.4 The sharp L^1 -Khintchine inequality: proof of Theorem K

The case $\frac{1}{2} < q \leq \frac{5}{8}$ of Theorem K follows from the case $1 - q$: $-\chi_i$ has the law of χ_i with q replaced by $1 - q$, and $|\sum_i \alpha_i \chi_i|$ and $2\sqrt{q(1-q)}$ do not change (the level-one form then follows by the equivalence at the end of the proof, which is valid for every q). In the rest of this subsection $\frac{3}{8} \leq q \leq \frac{1}{2}$ and χ denotes a random variable with $\Pr[\chi = u] = q$ and $\Pr[\chi = -v] = 1 - q$, where $u = \sqrt{(1-q)/q}$ and $v = \sqrt{q/(1-q)}$; $\mathbb{E}\chi = 0$ and $\mathbb{E}\chi^2 = 1$. Put (local notation)

$$\tau = \sqrt{q(1-q)}, \quad \theta = 1 - 2q, \quad \kappa = \frac{1-q}{4q}, \quad \kappa' = \frac{q}{4(1-q)}.$$

Then

$$qu = (1-q)v = \tau, \quad \tau v = q, \quad \tau u = 1 - q, \quad u + v = \frac{1}{\tau}, \quad \mathbb{E}|\chi| = 2\tau, \quad 4\kappa'\tau^2 = q^2.$$

The proof uses the following constants; here $a_B = \kappa/(1-\kappa) = \frac{1-q}{5q-1}$, $a_A \in (0, 1)$ is a threshold with

$$\sqrt{2/\pi} + a_A C_q \leq 2\tau, \quad C_q = \frac{q^2 + (1-q)^2}{2\tau} = \frac{1-2\tau^2}{2\tau}, \quad (11)$$

and p_q is the positive integer multiple of $2\theta s^3 + (\theta^2 + 2\kappa' - 1)s^2 + 2\theta\kappa's + \kappa'^2$ with coprime integer coefficients. The table gives them for three values of q ; for the other $q \in [\frac{3}{8}, \frac{1}{2}]$ see the end of the proof.

q	τ^2	$(\theta, \kappa, \kappa')$	a_A	a_B	$p_q(s)$	$[s_-, s_+]$
$\frac{3}{8}$	$\frac{15}{64}$	$(\frac{1}{4}, \frac{5}{12}, \frac{3}{20})$	$\frac{3}{10}$	$\frac{5}{7}$	$200s^3 - 255s^2 + 30s + 9$	$[\frac{31}{100}, \frac{21}{20}]$
$\frac{2}{5}$	$\frac{6}{25}$	$(\frac{1}{5}, \frac{3}{8}, \frac{1}{6})$	$\frac{1}{3}$	$\frac{3}{5}$	$360s^3 - 564s^2 + 60s + 25$	$[\frac{7}{20}, 1]$
$\frac{5}{12}$	$\frac{35}{144}$	$(\frac{1}{6}, \frac{7}{20}, \frac{5}{28})$	$\frac{7}{20}$	$\frac{7}{13}$	$2352s^3 - 4340s^2 + 420s + 225$	$[\frac{9}{25}, 1]$

For the level-one weight on a biased cube we use the basis of the previous subsection: if the coordinates of $x \in \{-1, 1\}^n$ are independent with $\Pr[x_i = 1] = \beta$, then $W^1(f) = \sum_i \hat{f}(\{i\})^2$ with $\hat{f}(\{i\}) = \mathbb{E}[f(x)\chi^{(\beta)}(x_i)]$; for $\beta = q$, $\chi^{(q)}(x_i)$ has the law of χ .

Lemma 9.4. *Let $a > 0$, $b \geq 0$, and let M be a bounded random variable independent of χ with $\mathbb{E}M = 0$. Then*

$$\mathbb{E}|a\chi + bM| = 2\tau a + 2\mathbb{E}R_a(bM), \quad R_a(y) = (1-q)(y - av)_+ + q(-y - au)_+.$$

Proof. For fixed y , using $|z| = z + 2(-z)_+$ for $z = au + y$ and $|z| = -z + 2z_+$ for $z = y - av$,

$$\mathbb{E}_\chi|a\chi + y| = q|au + y| + (1-q)|y - av| = qau + (1-q)av + (2q-1)y + 2R_a(y) = 2\tau a + (2q-1)y + 2R_a(y).$$

Put $y = bM$ and take expectations, using $\mathbb{E}M = 0$. □

Lemma 9.5 (Mills ratio). *Let $m(w) = e^{w^2/2} \int_w^\infty e^{-x^2/2} dx$. Then $\frac{w}{1+w^2} < m(w) < \frac{w^2+2}{w^3+3w}$ for all $w > 0$.*

Proof. Let $A(w) = \int_w^\infty e^{-x^2/2} dx - \frac{w}{1+w^2} e^{-w^2/2}$ and $B(w) = \frac{w^2+2}{w^3+3w} e^{-w^2/2} - \int_w^\infty e^{-x^2/2} dx$. Both tend to 0 as $w \rightarrow \infty$, and a direct computation gives $A'(w) = -\frac{2}{(1+w^2)^2} e^{-w^2/2} < 0$ and $B'(w) = -\frac{6}{w^2(w^2+3)^2} e^{-w^2/2} < 0$. Hence $A, B > 0$ on $(0, \infty)$. □

Lemma 9.6 (Stein function for $|x|$). *Let $f(w) = \sqrt{2/\pi} m(w) - 1$ for $w \geq 0$ and $f(w) = -f(-w)$ for $w < 0$. Then f is C^1 , f and f' are bounded, f' is 1-Lipschitz, and $f'(w) - wf(w) = |w| - \sqrt{2/\pi}$ for all $w \in \mathbb{R}$.*

Proof. We have $m' = wm - 1$ and $m(0) = \sqrt{\pi/2}$, so $f(0) = 0$ and the odd extension is continuous. For $w > 0$, $f'(w) = \sqrt{2/\pi}(wm(w) - 1)$ and $f''(w) = \sqrt{2/\pi}k(w)$ with $k(w) = (1 + w^2)m(w) - w$. The function f' is even, and $f'(0\pm) = -\sqrt{2/\pi}$, so $f \in C^1$. By Lemma 9.5, $0 < wm(w) < \frac{w^2+2}{w^2+3} < 1$, so $f' \in [-\sqrt{2/\pi}, 0]$, and m is decreasing, so $f \in (-1, 0]$ on $[0, \infty)$. For $w > 0$, $f'(w) - wf(w) = \sqrt{2/\pi}(wm - 1) - \sqrt{2/\pi}wm + w = w - \sqrt{2/\pi}$; for $w < 0$ the equation follows from $f'(w) - wf(w) = f'(-w) - (-w)f(-w)$, and at $w = 0$ it reads $f'(0) = -\sqrt{2/\pi}$. Finally, $k > 0$ on $(0, \infty)$ by the lower bound of Lemma 9.5, and $k'(w) = 2wm + (1 + w^2)(wm - 1) - 1 = (w^3 + 3w)m(w) - (w^2 + 2) < 0$ by the upper bound, so $0 < k(w) < k(0) = \sqrt{\pi/2}$ and $0 < f'' < 1$ on $(0, \infty)$. As f'' is odd, $|f''| < 1$ on $\mathbb{R} \setminus \{0\}$, and since f' is continuous it is 1-Lipschitz. $\square$

Lemma 9.7 (zero-bias coupling). *Let U be uniform on $[-v, u]$. (i) For every $\alpha \in \mathbb{R}$ and every absolutely continuous $G: \mathbb{R} \rightarrow \mathbb{R}$ with bounded G' , $\mathbb{E}[\chi G(\alpha\chi)] = \alpha \mathbb{E}[G'(\alpha U)]$. (ii) There is a coupling of χ and U with $\mathbb{E}|\chi - U| = \frac{q^2 + (1-q)^2}{2\tau}$.*

Proof. (i) $\mathbb{E}[\chi G(\alpha\chi)] = quG(\alpha u) - (1-q)vG(-\alpha v) = \tau(G(\alpha u) - G(-\alpha v)) = \tau \int_{-v}^u \alpha G'(\alpha s) ds = \alpha \mathbb{E}G'(\alpha U)$, since $\tau(u+v) = 1$. (ii) Let V be uniform on $[0, 1]$, $U = -v + V/\tau$, and $\chi = -v$ if $V < 1-q$, $\chi = u$ otherwise. Then $\mathbb{E}|\chi - U| = \int_0^{1-q} \frac{V}{\tau} dV + \int_{1-q}^1 \frac{1-V}{\tau} dV = \frac{(1-q)^2 + q^2}{2\tau}$. $\square$

Proof of Theorem K. We prove $\mathbb{E}|L| \leq 2\tau$ for $L = \sum_{i=1}^n \alpha_i \chi_i$, $\sum_i \alpha_i^2 = 1$, by induction on n ; for $n = 1$, $\mathbb{E}|L| = \mathbb{E}|\chi| = 2\tau$. Let $n \geq 2$. Relabelling and replacing α by $-\alpha$ (which does not change $|L|$), we may assume that $a := \alpha_1 = \max_i |\alpha_i| > 0$. Let $b = \sqrt{1-a^2}$ and $M = b^{-1} \sum_{i \geq 2} \alpha_i \chi_i$ if $b > 0$, $M = 0$ if $b = 0$. Then $L = a\chi_1 + bM$, M is independent of χ_1 , $\mathbb{E}M = 0$, $\mathbb{E}M^2 \leq 1$, and, if $b > 0$, $\mathbb{E}|M| \leq 2\tau$ by the induction hypothesis (M is a unit combination of $n-1$ copies of χ).

Case A: $a \leq a_A$. Let f be as in Lemma 9.6. Let I be a random index with $\Pr[I = i] = \alpha_i^2$, and for each i let U_i be coupled with χ_i as in Lemma 9.7(ii), using auxiliary randomness independent of everything else; put $L^* = L - \alpha_I \chi_I + \alpha_I U_I$. For each i , $L_i = L - \alpha_i \chi_i$ is independent of (χ_i, U_i) , so Lemma 9.7(i), applied conditionally on L_i with $G(y) = f(L_i + y)$, gives

$$\mathbb{E}[Lf(L)] = \sum_i \alpha_i \mathbb{E}[\chi_i f(L_i + \alpha_i \chi_i)] = \sum_i \alpha_i^2 \mathbb{E}[f'(L_i + \alpha_i U_i)] = \mathbb{E}f'(L^*).$$

By Lemma 9.6, with $Z \sim N(0, 1)$ and $\mathbb{E}|Z| = \sqrt{2/\pi}$,

$$\begin{aligned} \mathbb{E}|L| - \sqrt{2/\pi} &= \mathbb{E}[f'(L) - Lf(L)] = \mathbb{E}[f'(L) - f'(L^*)] \leq \mathbb{E}|L - L^*| \\ &= \sum_i \alpha_i^2 |\alpha_i| \mathbb{E}|\chi_i - U_i| \leq a C_q, \end{aligned}$$

so $\mathbb{E}|L| \leq \sqrt{2/\pi} + a_A C_q \leq 2\tau$ by (11). For $q = \frac{3}{8}, \frac{2}{5}, \frac{5}{12}$, (11) holds with the tabulated a_A : C_q equals $\frac{17}{8\sqrt{15}} < 0.5487$, $\frac{13}{10\sqrt{6}} < 0.5308$ and $\frac{37}{12\sqrt{35}} < 0.5212$, so $a_A C_q$ is less than 0.1647, 0.1770 and 0.1825. With $\sqrt{2/\pi} < 0.7979$ the left side of (11) is less than 0.9626, 0.9749, 0.9804 respectively, while 2τ equals $\frac{\sqrt{15}}{4} > 0.9682$, $\frac{2\sqrt{6}}{5} > 0.9797$ and $\frac{\sqrt{35}}{6} > 0.9860$.

Case B: $a \geq a_B$. For $t > 0$ and $y \in \mathbb{R}$, $(y-t)_+ \leq \frac{y^2}{4t}$, because $\frac{y^2}{4t} - (y-t) = \frac{(y-2t)^2}{4t}$. At most one of the two terms of $R_a(y)$ is nonzero, and $\frac{1-q}{4av} = \frac{(1-q)u}{4a} = \frac{\kappa\tau}{a} \geq \frac{\kappa'\tau}{a} = \frac{qv}{4a} = \frac{q}{4au}$ (as $q \leq \frac{1}{2}$), so

$R_a(y) \leq \frac{\kappa\tau}{a} y^2$. By Lemma 9.4,

$$\mathbb{E}|L| \leq 2\tau a + \frac{2\kappa\tau}{a} b^2 \mathbb{E}M^2 \leq 2\tau \left(a + \frac{\kappa(1-a^2)}{a} \right) = 2\tau \left(1 + \frac{(a-1)((1-\kappa)a - \kappa)}{a} \right) \leq 2\tau,$$

since $a \leq 1$ and $(1-\kappa)a \geq (1-\kappa)a_B = \kappa$ (note $\kappa < 1$).

Case C: $a_A < a < a_B$. Here $b > 0$. We claim that for all $y \in \mathbb{R}$

$$R_a(y) \leq \frac{\kappa'\tau}{a} y^2 + \theta y_+. \quad (12)$$

Let $c = \kappa'\tau/a$, so that $4ca\tau = 4\kappa'\tau^2 = q^2$. For $-au \leq y \leq av$ the left side vanishes. For $y > av$ the difference of the two sides is $cy^2 + (\theta - 1 + q)y + (1-q)av = cy^2 - qy + a\tau$, whose discriminant is $q^2 - 4ca\tau = 0$. For $y < -au$ it is $cy^2 + qy + qau = cy^2 + qy + a\tau$, with the same discriminant. So both are $c > 0$ times squares, proving (12). Since $\mathbb{E}(bM)_+ = \frac{b}{2}\mathbb{E}|M| \leq b\tau$ (as $\mathbb{E}M = 0$), Lemma 9.4 and (12) give

$$\mathbb{E}|L| \leq 2\tau a + 2cb^2 + 2\theta b\tau = 2\tau \psi(a), \quad \psi(a) = a + \frac{\kappa'(1-a^2)}{a} + \theta\sqrt{1-a^2}.$$

It remains to show $\psi(a) \leq 1$. Put $s = a/b$. Dividing by b , $\psi(a) \leq 1$ is equivalent to $s + \frac{\kappa'}{s} + \theta \leq \sqrt{1+s^2}$, and, as both sides are positive, to $(s + \theta + \frac{\kappa'}{s})^2 \leq 1 + s^2$, i.e. (multiplying by s^2) to $2\theta s^3 + (\theta^2 + 2\kappa' - 1)s^2 + 2\theta\kappa's + \kappa'^2 \leq 0$, i.e. (multiplying by $16(1-q)^2$) to

$$P(q, s) := 32(1-q)^2(1-2q)s^3 + 8q(1-q)(1-8(1-q)^2)s^2 + 8q(1-q)(1-2q)s + q^2 \leq 0, \quad (13)$$

i.e. to $p_q(s) \leq 0$ for the tabulated q . The values of p_q at $-1, 0, s_-, s_+, 2$ are

$$\begin{aligned} & -476, 9, -\frac{2473}{10000}, -\frac{729}{80}, 649 & (q = \frac{3}{8}), \\ & -959, 25, -\frac{1531}{200}, -119, 769 & (q = \frac{2}{5}), \\ & -6887, 225, -\frac{1195767}{15625}, -1343, 2521 & (q = \frac{5}{12}). \end{aligned}$$

So the cubic p_q has a root in each of $(-1, 0)$, $(0, s_-)$ and $(s_+, 2)$, hence none in $[s_-, s_+]$, where it is therefore negative. Finally $s = a/\sqrt{1-a^2}$ is increasing in a , and $[s(a_A), s(a_B)] \subseteq [s_-, s_+]$: $s(\frac{3}{10}) = \frac{3}{\sqrt{91}} > 0.31$ and $s(\frac{5}{7}) = \frac{5}{\sqrt{24}} < 1.05$; $s(\frac{1}{3}) = \frac{1}{\sqrt{8}} > 0.35$ and $s(\frac{3}{5}) = \frac{3}{4}$; $s(\frac{7}{20}) = \frac{7}{\sqrt{351}} > 0.36$ and $s(\frac{7}{13}) = \frac{7}{\sqrt{120}} < 1$. Hence $\mathbb{E}|L| \leq 2\tau$, completing the induction for these three values of q .

Other values of $q \in [\frac{3}{8}, \frac{1}{2}]$. The argument above uses about q only that $\frac{1}{3} \leq q \leq \frac{1}{2}$ (in Case B: $\kappa < 1$, $\kappa \geq \kappa'$ and $a_B \leq 1$), the inequality (11) in Case A, and (13) for $s(a_A) < s < s(a_B)$ in Case C, where $s(a) = a/\sqrt{1-a^2}$; the threshold a_A may depend on q . Cover $[\frac{3}{8}, \frac{1}{2}]$ by the 64 intervals $[q_0, q_1] = [\frac{3}{8} + \frac{j}{512}, \frac{3}{8} + \frac{j+1}{512}]$, $0 \leq j \leq 63$, and fix one threshold a_A for each interval. Since τ is increasing in q on $[0, \frac{1}{2}]$ and $C_q = \frac{1}{2\tau} - \tau$ is decreasing in τ , (11) holds for all $q \in [q_0, q_1]$ as soon as $\frac{79789}{10^5} + a_A \frac{1-2\tau_0^2}{2\tau_0} \leq 2\tau_0$ for a rational $\tau_0 \leq \sqrt{q_0(1-q_0)}$, because $\sqrt{2/\pi} < \frac{79789}{10^5}$ (as $\pi > 3.14159$); we take for a_A the largest multiple of 10^{-4} with this property (for example $a_A = \frac{194}{625}$ for $j = 0$ and $a_A = \frac{4041}{10^4}$ for $j = 63$). As $a_B = \frac{1-q}{5q-1}$ is decreasing in q , Case C on $[q_0, q_1]$ only involves $a \in (a_A, a_B(q_0))$, which is empty for the 16 intervals with $j \geq 48$ (i.e. $q_0 \geq \frac{15}{32}$). For each of the other 48 intervals take rationals $s_{\text{lo}} \leq s(a_A)$ and $s_{\text{hi}} \geq s(a_B(q_0))$. On a box $[q', q''] \times [s', s''] \subseteq [\frac{3}{8}, \frac{1}{2}] \times [0, \infty)$,

$$P(q, s) \leq 32(1-q')^2(1-2q')s'''^3 - 8q'(1-q')(8(1-q'')^2 - 1)s'^2 + 8q''(1-q'')(1-2q')s'' + q''^2,$$

since the coefficients of s^3 , s and 1 in P are nonnegative, that of s^2 is $-8q(1-q)(8(1-q)^2-1) \leq -8q(1-q) < 0$, and each factor is monotone in q . Bisectioning $[q_0, q_1] \times [s_{\text{lo}}, s_{\text{hi}}]$ until this bound is negative on every box proves $P < 0$ there; 249 boxes suffice in total (maximal bisection depth 6). All arithmetic is exact rational arithmetic (`khintchine_interval.py`). This proves Theorem K for $\frac{3}{8} \leq q \leq \frac{1}{2}$, and hence, by the symmetry noted at the start of this subsection, for $\frac{3}{8} \leq q \leq \frac{5}{8}$.

Equality for $\alpha = \pm e_1$ is clear. For the equivalent form: if $f: \{-1, 1\}^n \rightarrow [0, 1]$ and $W^1(f) > 0$, let $\alpha_i = \hat{f}(\{i\})/\sqrt{W^1(f)}$ and $L = \sum_i \alpha_i \chi^{(q)}(x_i)$. Then $\sqrt{W^1(f)} = \mathbb{E}[fL] = \mathbb{E}[(f - \frac{1}{2})L] \leq \frac{1}{2}\mathbb{E}|L| \leq \tau$, i.e. $W^1(f) \leq \tau^2 = q(1-q)$; for $f = \mathbf{1}\{x_1 = 1\}$, $\hat{f}(\{1\}) = qu = \tau$. Conversely, given α take $f = \mathbf{1}\{L > 0\}$: then $\mathbb{E}|L| = 2\mathbb{E}[fL] = 2\sum_i \alpha_i \hat{f}(\{i\}) \leq 2\sqrt{W^1(f)} \leq 2\tau$. $\square$

Remark 9.8. (a) The margins are small but explicit: $\max \psi$ over $[a_A, a_B]$ is below 0.9935, 0.9664 and 0.9539 for $q = \frac{3}{8}, \frac{2}{5}, \frac{5}{12}$, and in Case A, $2\tau - 0.7979 - a_A \frac{q^2 + (1-q)^2}{2\tau} \geq 0.0049$ (the rounded bounds displayed in the proof only show 0.0048 for $q = \frac{2}{5}$); both statements are checked exactly. Exact checks of the table, of the identities used for general q and of the case computations are in `khintchine_q_check.py`, which also verifies (12) exactly (through the two discriminants); floating-point checks of (12) and of ψ are in `khintchine_analytic_check.py` ($q = \frac{3}{8}$). Both scripts evaluate $\mathbb{E}|L|$ (exact enumeration over $\{u, -v\}^n$) for random unit vectors, which are near-dictator sanity checks, and `khintchine_q_check.py` adds a local search started away from dictators. For $q = \frac{3}{8}$, an independent computer-assisted proof of Case C, by exact rational dual certificates of the moment problem “maximise $\mathbb{E}R_a(bM)$ given $\mathbb{E}M = 0$, $\mathbb{E}M^2 = 1$, $\mathbb{E}|M| \leq 2\tau$ ” on 560 cells, is `khintchine_certificates.py`. Its numerically optimal certificates have the shape of (12) (no constant term, the two outer pieces tangent to R_a); (12) fixes the coefficient of $|y|$ at $\frac{\theta}{2}$, which gives a closed form at the price of a smaller margin. (b) For small a one must use more about M than a few moments: for $q = \frac{3}{8}$ the moment problem “maximise $\mathbb{E}R_a(bM)$ subject to $\mathbb{E}M = 0$, $\mathbb{E}M^2 = 1$, $\mathbb{E}|M| \leq 2\tau$ ” has value larger than $\tau(1-a)$ at $a = 0.05$ and $a = 0.1$ (`khintchine_moment_lp.py`; discretised linear programme, floating point), which is why Case A uses the zero-bias coupling of the whole sum. Case B is nonempty only if $a_B \leq 1$, i.e. $q \geq \frac{1}{3}$, and Case A only if $2\sqrt{q(1-q)} > \sqrt{2/\pi}$. With the best possible threshold in Case A, a floating-point scan (`khintchine_q_check.py`, grid of step 0.005 in q) indicates that Case C fails for $q \leq 0.37$ ($\max \psi > 1$), so the method does not reach much below $\frac{3}{8}$; for $q \geq 0.47$ Cases A and B overlap, in line with the 16 intervals on which Case C is empty in the proof. (c) The statement fails for every $q < 1 - \frac{1}{\sqrt{2}} \approx 0.2929$ (and, by symmetry, for $q > \frac{1}{\sqrt{2}}$), by the two-equal-coefficient example of Stone [16, §4.2]: for $\alpha = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ the sum L takes the values $\sqrt{2}u, \frac{u-v}{\sqrt{2}}, -\sqrt{2}v$ with probabilities $q^2, 2q(1-q), (1-q)^2$, so $\mathbb{E}|L| = \sqrt{2}(q+1-2q+1-q)\tau = \sqrt{2}(1-q) \cdot 2\tau$. Numerically this is the only obstruction: maximising $\mathbb{E}|L|/(2\tau)$ over unit vectors with $n \leq 7$ (exact enumeration, local search from random starts; `khintchine_threshold.py`) gives exactly 1, attained at $\pm e_i$, for every q of a grid in $[0.2929, 0.375]$, and the value $\sqrt{2}(1-q)$ for $q = 0.28, 0.29$. We conjecture that Theorem K holds exactly for $1 - \frac{1}{\sqrt{2}} \leq q \leq \frac{1}{\sqrt{2}}$; at $q = 1 - \frac{1}{\sqrt{2}}$ both e_1 and $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ would be extremal. (d) Literature. The ratio $R(q) = \sup \mathbb{E}|L|/(\mathbb{E}L^2)^{1/2}$, over all n and all $\alpha \neq 0$, for the same normalised q -biased bits appears in work on the L^1 -Poincaré inequality on the Hamming cube. Ivanisvili, Li, van Handel and Volberg [9] showed that $R(q) < 1$ for q in a neighbourhood of $\frac{3}{4}$ and used this to prove that the constant $\frac{\pi}{2}$ in that inequality is not sharp; a quantitative bound $R(q) \leq 1 - \frac{1}{480} \min\{4q^3(1-q)^3, (2q-1)^4 q(1-q)\}$, due to van Handel [17], is quoted by Ivanisvili and Stone [10], who improve it and discuss the exact value of $R(q)$; see also Stone’s thesis [16], whose §4.2 contains the two-coefficient example of (c). Theorem K gives the exact value $R(q) = 2\sqrt{q(1-q)}$, attained at $n = 1$, for $q \in \{\frac{3}{8}, \frac{2}{5}, \frac{5}{12}\}$ and the mirror values $\frac{5}{8}, \frac{3}{5}, \frac{7}{12}$ (by hand), and for all $q \in [\frac{3}{8}, \frac{5}{8}]$ with the computer-assisted

interval verification. Case A uses a zero-bias Stein bound, as does Lei Yu [11] for the sharp small-coefficient L^1 -Khintchine inequality for unbiased bits. For nonsymmetric Khinchine–Kahane inequalities with constants that are optimal up to a universal factor see Oleszkiewicz [15], and for zero-bias couplings Goldstein–Reinert [7]. (e) Code. The scripts cited in this remark are in the repository, with their outputs: `numerics/khintchine_certificates.py` (output `numerics/outputs/khintchine_certificates.txt`; the independent proof of Case C), `numerics/khintchine_threshold.py` (output `khintchine_threshold.txt`; part (c)), `khintchine_q_check.py`, `khintchine_analytic_check.py` and `khintchine_moment_lp.py`.

9.5 Proof of Proposition 9.1

Fix k, ν, F and let P, p, W_+ be as above. By Theorem K with $q = \frac{3}{8}$ (in its level-one form, applied to $\bar{P}$), $W_+ \leq \frac{15}{64}$; also $W_+ \leq \text{Var}P \leq p - p^2$. By Lemmas 9.2 and 9.3, and since $\frac{43}{15} - a^* > 0$,

$$2\Gamma_\nu(F) + 1 \leq 2p - 3p^2 + a^*(p - p^2) + \left(\frac{43}{15} - a^*\right) \min\left(p - p^2, \frac{15}{64}\right).$$

If $p \leq \frac{3}{8}$ or $p \geq \frac{5}{11}$, use the first term of the minimum: the right-hand side is at most $2p - 3p^2 + \frac{43}{15}(p - p^2) = 1 - \frac{(8p-3)(11p-5)}{15} \leq 1$. If $\frac{3}{8} \leq p \leq \frac{5}{11}$, use $\frac{15}{64}$: the right-hand side is $G(p) = 2p - 3p^2 + a^*(p - p^2) + \left(\frac{43}{15} - a^*\right)\frac{15}{64}$, a concave quadratic with $G(\frac{3}{8}) = \frac{3}{4} - \frac{27}{64} + \frac{43}{64} = 1$ and $G'(\frac{3}{8}) = \frac{a^*-1}{4} < 0$, so $G(p) \leq 1$ for $p \geq \frac{3}{8}$. In both cases $\Gamma_\nu(F) \leq 0$. $\square$

Remark 9.9. Every inequality in the proof is an equality for $k = 1, \nu = (+1), F(z) = z$, where $\Gamma_+ = 0$. Then $P = \mathbf{1}\{z = 1\}$, $p = \frac{3}{8}$, $P_1P_2 = 0$ almost surely (as $\pi_+(++\cdot) = 0$), $W_+ = \text{Var}P = \frac{15}{64}$, and the extremal case of Theorem K is attained. This is why neither bound on W_+ suffices alone: with $W_+ \leq p - p^2$ the right-hand side exceeds 1 for $\frac{3}{8} < p < \frac{5}{11}$, and with $W_+ \leq \frac{15}{64}$ it exceeds 1 just below $p = \frac{3}{8}$ ($\max G = 1 + 4.7 \cdot 10^{-6}$, at $p \approx 0.3739$).

9.6 A second function: g_1

With inputs $(w_1, w_2, w_3, u_1, u_2)$ as before, let

$$g_1(w, u) = +1 \iff (u_1, u_2) \in A'_w, \quad \text{where} \quad (14)$$

$$\begin{aligned} A'_{+++} &= A'_{++-} = \emptyset, & A'_{+-+} &= \{(+, +)\}, & A'_{+--} &= \{\pm 1\}^2 \setminus \{(+, +)\}, \\ A'_{-++} &= \{(+, +), (-, -)\}, & A'_{-+-} &= \{(+, -), (-, +)\}, & A'_{--+} &= \{(+, -)\}, & A'_{---} &= \{(-, +)\}. \end{aligned}$$

Then $h(w) = \frac{1}{2}|A'_w| - 1$, that is, in the order $++++, ++-+, \dots, ----$,

$$h = \left(-1, -1, -\frac{1}{2}, \frac{1}{2}, 0, 0, -\frac{1}{2}, -\frac{1}{2}\right), \quad \mu = -\frac{3}{8}, \quad \sigma = -1, \quad \text{triv}_{g_1} = \frac{11}{16}.$$

A direct count gives $\hat{g}_1(\{w_j\}) = \hat{h}(\{j\}) = -\frac{1}{8}$ for $j = 1, 2, 3$ and

$$\hat{g}_1(\{u_i\}) = 2 \cdot 2^{-5} \sum_w \sum_{(u_1, u_2) \in A'_w} u_i = 0 \quad (i = 1, 2)$$

(over $w = +-+, +--, -++ , -+-, ---$ the inner sums are 1, $-1, 0, 0, 1, -1$ for $i = 1$ and 1, $-1, 0, 0, -1, 1$ for $i = 2$). So $J = \{1, 2, 3\}$ is a single class. The function g_1 depends on u_1 and u_2 (as $g_1(-, +, +, u) = u_1u_2$) and on each w_j , it is not $\pm$ a parity and $\mu \neq 0$, so Theorems B–D apply. Its h belongs to the list `hard288`, in an orbit of size 24 (one of the five one-class orbits of §11). The row laws are

$$\pi_+ = \frac{1+h}{5} = \left(0, 0, \frac{1}{10}, \frac{3}{10}, \frac{1}{5}, \frac{1}{5}, \frac{1}{10}, \frac{1}{10}\right), \quad \pi_- = \frac{1-h}{11} = \left(\frac{2}{11}, \frac{2}{11}, \frac{3}{22}, \frac{1}{22}, \frac{1}{11}, \frac{1}{11}, \frac{3}{22}, \frac{3}{22}\right),$$

and each coordinate w_j equals $+1$ with probability $\beta_+ = \frac{2}{5}$ under π_+ and $\beta_- = \frac{6}{11}$ under π_- . With k, ν, F, C, X, Y, Z and $\Gamma_\nu(F) = \mathbb{E} h(F(X), F(Y), F(Z))$ as in §9, and with the basis χ_S of $L^2(\Omega_\nu)$ now built from $\beta_\pm = \frac{2}{5}, \frac{6}{11}$:

Proposition 9.10. $\Gamma_\nu(F) \leq 0$ for all $k \geq 0, \nu \in \{-1, 1\}^k$ and $F: \{-1, 1\}^k \rightarrow [-1, 1]$. Consequently $s_{g_1}^L = s_{g_1} = \text{triv}_{g_1} = \frac{1}{16} < s_{g_1}^U$; the same holds for $g_{1,m}(w, u_1, \dots, u_{m-3}) = g_1(w, u_1, u_2 \cdots u_{m-3})$ for every $m \geq 5$, and for every g as in Theorem B whose h lies in the orbit of h_{g_1} under the symmetries of Lemma 11.2.

Proof. The consequences follow from the first statement exactly as Theorem E and Corollary E' follow from Proposition 9.1 (here $\sigma h(+, -, -) = -\frac{1}{2} < 0$, and the product $u_2 \cdots u_{m-3}$ preserves h, μ and the vanishing of the u -coefficients), together with Lemma 11.2 for the orbit; note that the criterion of Theorem C(ii) is invariant under the same symmetries.

Let $P = \frac{1+F}{2}$, $P_1 = P(X)$, $P_2 = P(Y)$, $P_3 = P(Z)$ and $p = \mathbb{E}P$. As in Lemma 9.2, both sides of

$$2h(y) = -1 + 2p_1 + p_2 - 4p_1p_2 - 2p_1p_3 + 2p_1p_2p_3 \quad (p_j = \frac{1+y_j}{2})$$

are multilinear and agree at the vertices: the right-hand side takes the values $-2, -2, -1, 1$ at $p = (1, 1, 1), (1, 1, 0), (1, 0, 1), (1, 0, 0)$ and $0, 0, -1, -1$ at $p = (0, 1, 1), (0, 1, 0), (0, 0, 1), (0, 0, 0)$. Since $P_1P_2P_3 \leq P_1P_2$ and X, Y have the same law,

$$2\Gamma_\nu(F) \leq -1 + 3p - 2\mathbb{E}[P_1P_2] - 2\mathbb{E}[P_1P_3].$$

The law of (w_1, w_2) on $(++, +-, -+, --)$ is $(0, \frac{2}{5}, \frac{2}{5}, \frac{1}{5})$ under π_+ and $(\frac{4}{11}, \frac{2}{11}, \frac{2}{11}, \frac{3}{11})$ under π_- ; the law of (w_1, w_3) is $(\frac{1}{10}, \frac{3}{10}, \frac{3}{10}, \frac{3}{10})$ under π_+ and $(\frac{7}{22}, \frac{5}{22}, \frac{5}{22}, \frac{5}{22})$ under π_- . With the formula for $\mathbb{E}[\chi^{(\beta)}(t)\chi^{(\beta)}(t')]$ from the proof of Lemma 9.3, the correlations are

$$\lambda_+^{12} = \frac{-\frac{3}{5} - \frac{1}{25}}{\frac{24}{25}} = -\frac{2}{3}, \quad \lambda_-^{12} = \frac{\frac{3}{11} - \frac{1}{121}}{\frac{120}{121}} = \frac{4}{15}, \quad \lambda_+^{13} = \frac{-\frac{1}{5} - \frac{1}{25}}{\frac{24}{25}} = -\frac{1}{4}, \quad \lambda_-^{13} = \frac{\frac{1}{11} - \frac{1}{121}}{\frac{120}{121}} = \frac{1}{12},$$

and, as in that proof, $\mathbb{E}[P_1P_l] = \sum_S \lambda_S^{1l} \hat{P}(S)^2$ for $l = 2, 3$. Hence

$$2\Gamma_\nu(F) \leq -1 + 3p - 4p^2 + \sum_{S \neq \emptyset} a_S \hat{P}(S)^2, \quad a_S = -2(\lambda_S^{12} + \lambda_S^{13}).$$

If $S = \{i\}$ then $a_S = 2(\frac{2}{3} + \frac{1}{4}) = \frac{11}{6}$ when $\nu_i = +1$ and $a_S = -2(\frac{4}{15} + \frac{1}{12}) = -\frac{7}{10}$ when $\nu_i = -1$. If $|S| = 2$, a_S is $-2(\frac{4}{9} + \frac{1}{16}) < 0$, or $2(\frac{8}{45} + \frac{1}{48}) = \frac{143}{360}$, or $-2(\frac{16}{225} + \frac{1}{144}) < 0$, according as S contains two, one or no indices with $\nu_i = +1$. If $|S| \geq 3$ then $a_S \leq 2((\frac{2}{3})^{|S|} + (\frac{1}{4})^{|S|}) \leq a_1^* := 2((\frac{2}{3})^3 + (\frac{1}{4})^3) = \frac{539}{864}$. So every coefficient other than the level-one coefficient of the $\nu = +1$ rows is at most a_1^* , and with W_+ as in §9

$$2\Gamma_\nu(F) \leq -1 + 3p - 4p^2 + a_1^*(p - p^2) + (\frac{11}{6} - a_1^*)W_+.$$

As in §9, W_+ is the level-one weight of a $[0, 1]$ -valued function on the $\frac{2}{5}$ -biased cube, so $W_+ \leq \frac{6}{25}$ by Theorem K with $q = \frac{2}{5}$; also $W_+ \leq \text{Var}P \leq p - p^2$. If $p \leq \frac{2}{5}$ or $p \geq \frac{3}{7}$, the bound $W_+ \leq p - p^2$ gives $2\Gamma_\nu(F) \leq -1 + 3p - 4p^2 + \frac{11}{6}(p - p^2) = -\frac{1}{6}(5p - 2)(7p - 3) \leq 0$. If $\frac{2}{5} \leq p \leq \frac{3}{7}$, the bound $W_+ \leq \frac{6}{25}$ gives $2\Gamma_\nu(F) \leq G_1(p) := -1 + 3p - 4p^2 + a_1^*(p - p^2) + (\frac{11}{6} - a_1^*)\frac{6}{25}$, a concave quadratic with $G_1(\frac{2}{5}) = 0$ and $G_1'(\frac{2}{5}) = \frac{a_1^* - 1}{5} < 0$, so $G_1(p) \leq 0$ for $p \geq \frac{2}{5}$. $\square$

As for g_0 , every inequality is tight for the dictator decoder ($k = 1, \nu = +1, F(z) = z$: $p = \frac{2}{5}, P_1P_2 = 0$ almost surely because $\pi_+(++\cdot) = 0$, and $W_+ = \text{Var}P = \frac{6}{25}$), which gives $\Gamma_+ = 0$. All numbers in this subsection, and the identity between the reduction bound and the spectral

formula on random decoders, are checked in `g1_check.py`. Two later proofs of Proposition 9.10 avoid Theorem K: one through the two-class orbit rep7 (Remark 11.11(b)), and a shorter one that keeps the first step above ($P_1 P_2 P_3 \leq P_1 P_2$) and replaces the spectral bound by a re-coupling of the rows (§11, paragraph “One-class designs”).

9.7 A level-one inequality for nonnegative functions: proof of Theorem L

Let $0 < q \leq \frac{1}{2}$, let $\{-1, 1\}^n$ carry the q -biased product measure, and use the basis χ_S built from $\chi^{(q)}$ as in §9, so that $\mathbb{E}f = \hat{f}(\emptyset)$, $\mathbb{E}f^2 = \sum_S \hat{f}(S)^2$ and $W^1(f) = \sum_{|S|=1} \hat{f}(S)^2$. Put

$$\beta = \frac{1-2q}{q}, \quad \rho_* = -\frac{q}{1-q}, \quad c = \frac{(1-q)^2}{q}.$$

Let (x, y) be a random pair of points of $\{-1, 1\}^n$ whose coordinate pairs (x_i, y_i) are independent, each with

$$\Pr[x_i = y_i = 1] = 0, \quad \Pr[x_i = 1, y_i = -1] = \Pr[x_i = -1, y_i = 1] = q, \quad \Pr[x_i = y_i = -1] = 1 - 2q$$

(nonnegative as $q \leq \frac{1}{2}$). Then x and y both have the q -biased law, and $\mathbb{E}[x_i y_i] = 1 - 4q$, so by the formula for $\mathbb{E}[\chi^{(\beta)}(t)\chi^{(\beta)}(t')]$ in the proof of Lemma 9.3, $\mathbb{E}[\chi^{(q)}(x_i)\chi^{(q)}(y_i)] = \frac{(1-4q)-(1-2q)^2}{1-(1-2q)^2} = -\frac{4q^2}{4q(1-q)} = \rho_*$. As in that proof, $\mathbb{E}[f(x)f(y)] = \sum_S \rho_*^{|S|} \hat{f}(S)^2$ for every f , and therefore

$$q \mathbb{E}f^2 + \beta(\mathbb{E}f)^2 - W^1(f) = c \mathbb{E}[f(x)f(y)] + \sum_S e_{|S|} \hat{f}(S)^2, \quad (15)$$

where e_k is the coefficient of $\hat{f}(S)^2$, $|S| = k$, on the left minus $c\rho_*^k$:

$$e_0 = q + \beta - c = 0, \quad e_1 = q - 1 - c\rho_* = q - 1 + (1 - q) = 0, \quad e_2 = q - c\rho_*^2 = q - q = 0,$$

$$e_k = q - c\rho_*^k = q \left(1 - (-1)^k \left(\frac{q}{1-q} \right)^{k-2} \right) \geq 0 \quad (k \geq 3).$$

If $f \geq 0$ then $f(x)f(y) \geq 0$, so both terms on the right of (15) are nonnegative, which proves Theorem L; for $f: \{-1, 1\}^n \rightarrow [0, 1]$ use $\mathbb{E}f^2 \leq \mathbb{E}f$, with equality iff f is $\{0, 1\}$ -valued. For $q < \frac{1}{2}$, $c > 0$, $e_k > 0$ for $k \geq 3$ (as $\frac{q}{1-q} < 1$), and the law of (x, y) gives positive mass to every pair whose sets of $+1$ coordinates are disjoint; so both terms vanish iff $\hat{f}(S) = 0$ for $|S| \geq 3$ and $f(x)f(y) = 0$ for all such pairs. For $q = \frac{1}{2}$, $y = -x$ almost surely, $e_k = 0$ for even k and $e_k = 2q > 0$ for odd $k \geq 3$, which gives the stated condition.

It remains to make the case $q < \frac{1}{2}$ explicit. Write f as a multilinear polynomial in $y_i = \mathbf{1}\{x_i = 1\}$; the condition $\hat{f}(S) = 0$ for $|S| \geq 3$ says that it has degree at most 2, $f = c_0 + \sum_i c_i y_i + \sum_{i < j} c_{ij} y_i y_j$. Let e_A be the point whose set of $+1$ coordinates is A , and assume the second condition. With $x = y = e_\emptyset$ it gives $c_0 = f(e_\emptyset) = 0$, and with $x = e_{\{i\}}$, $y = e_{\{j\}}$ it gives $c_i c_j = 0$ for $i \neq j$; also $c_i = f(e_{\{i\}}) \geq 0$. If $c_1 > 0$, say, then $f(x) = 0$ whenever $x_1 = -1$ (take $y = e_{\{1\}}$), so all terms without y_1 vanish and $f = y_1(c_1 + \sum_{j \geq 2} c_{1j} y_j)$, which is the first form. If all $c_i = 0$, then $c_{ij} = f(e_{\{i,j\}}) \geq 0$, and the pairs $\{i, j\}$ with $c_{ij} > 0$ pairwise meet (take $x = e_{\{i,j\}}$, $y = e_{\{k,l\}}$); so they form a star, which is the first form with h linear, or lie in a triangle, which is the second form. Conversely, both forms have degree at most 2, and $f(x)f(y) \neq 0$ needs $x_1 = y_1 = 1$ in the first form and at least two $+1$ coordinates among the first three in both x and y in the second, so the sets of $+1$ coordinates of x and y meet. Finally, a $\{0, 1\}$ -valued h of degree at most 1 is

constant or equal to $\mathbf{1}\{x_j = 1\}$ or $\mathbf{1}\{x_j = -1\}$ for some j , and a $\{0, 1\}$ -valued f of the second form has at most one of a, b, c nonzero (its value at $e_{\{1,2,3\}}$ is $a + b + c$), which gives the $\{0, 1\}$ -valued equality cases. $\square$

The operator $f \mapsto \mathbb{E}[f(y) \mid x = \cdot]$ is the one Friedgut uses for the biased Erdős–Ko–Rado theorem [5]. The pair (x, y) is supported on pairs with disjoint sets of $+1$ coordinates. So if $A \subseteq \{-1, 1\}^n$ is intersecting (the sets of $+1$ coordinates of any two, not necessarily distinct, members of A meet), then $\mathbb{E}[\mathbf{1}_A(x)\mathbf{1}_A(y)] = 0$, and (15) reads $q\mu(A) + \beta\mu(A)^2 - W^1(\mathbf{1}_A) = \sum_{|S| \geq 3} e_{|S|} \hat{\mathbf{1}}_A(S)^2$.

Sharpness and small means. Let $q < \frac{1}{2}$. The first inequality is homogeneous of degree 2, and an equality case $f \neq 0$ has $r = (\mathbb{E}f)^2/\mathbb{E}f^2 \in [q^2, q]$. For the first form, $r = q(\mathbb{E}h)^2/\mathbb{E}h^2$, and $h = d_0 + \sum_j d_j y_j \neq 0$ with $h \geq 0$ has $(\mathbb{E}h)^2 \leq \mathbb{E}h^2 \leq (\max h)\mathbb{E}h$ and $\mathbb{E}h - q \max h = (1-q)d_0 - q \sum_j d_j^- \geq (1-2q) \sum_j d_j^- \geq 0$, where $d_j^- = \max(-d_j, 0)$ and $d_0 \geq \sum_j d_j^-$ because $h \geq 0$; so $(\mathbb{E}h)^2/\mathbb{E}h^2 \in [q, 1]$. For the second form, $r = q^2(a+b+c)^2/(a^2+b^2+c^2+2q(ab+ac+bc))$, and $q^2 \leq r \leq q$ as $q \leq 1$. Every $r \in [q^2, q]$ occurs, for $f = \mathbf{1}\{x_1 = 1\}(\alpha + \beta'\mathbf{1}\{x_2 = 1\})$ with $\alpha, \beta' \geq 0$. The second inequality is linear in p and far from sharp for small p . For the AND of t coordinates, $f = \mathbf{1}\{x_1 = \dots = x_t = 1\}$, we have $p = q^t$ and $\hat{f}(\{i\}) = p\sqrt{(1-q)/q}$ for $i \leq t$, so $W^1(f) = t \frac{1-q}{q} p^2$ and

$$\frac{W^1(f)}{qp + \beta p^2} = \frac{t(1-q)q^{t-2}}{1 + (1-2q)q^{t-2}}.$$

This is 1 for $t = 1, 2$ and $\frac{3q}{1+2q}$ for $t = 3$. It is less than 1 for every $t \geq 3$: with $s = t - 2$ the condition reads $q^s(s+1-sq) < 1$, and $q \mapsto q^s(s+1-sq)$ has derivative $s(s+1)q^{s-1}(1-q) > 0$ on $(0, 1)$ and value 1 at $q = 1$. It tends to 0 as $t \rightarrow \infty$. The correct order for small means is $p^2 \ln \frac{1}{p}$: every $f: \{-1, 1\}^n \rightarrow [0, 1]$ with $\mathbb{E}f = p \in (0, 1)$ satisfies

$$W^1(f) \leq \frac{p^2}{2q(1-q)} \left(\ln \frac{1}{p} + 1 + \frac{1}{4 \ln(1/p)} \right), \quad (16)$$

while for $p = q^t$ the AND of t coordinates has $W^1 = \frac{1-q}{q \ln(1/q)} p^2 \ln \frac{1}{p}$. This is the biased version of the level-one inequality [14], with the usual proof. Let $\ell = \sum_i \hat{f}(\{i\}) \chi_{\{i\}}$, so that $W := W^1(f) = \mathbb{E}\ell^2 = \mathbb{E}[f\ell]$, and assume $W > 0$. The summands of $Y = \ell/\sqrt{W}$ are independent with mean 0, and the i -th takes values in an interval of length $|\hat{f}(\{i\})|/(\tau\sqrt{W})$, $\tau = \sqrt{q(1-q)}$, because $\chi^{(q)}$ takes the values $\sqrt{(1-q)/q}$ and $-\sqrt{q/(1-q)}$, whose difference is $1/\tau$. So Hoeffding's inequality gives $\Pr[Y \geq y] \leq e^{-2\tau^2 y^2}$ for $y \geq 0$, and as $0 \leq f \leq 1$,

$$\begin{aligned} \sqrt{W} = \mathbb{E}[fY] &\leq \int_0^\infty \mathbb{E}[f \mathbf{1}\{Y \geq y\}] dy \leq \int_0^\infty \min(p, e^{-2\tau^2 y^2}) dy \\ &\leq py_0 + \frac{p}{4\tau^2 y_0}, \quad y_0 = \sqrt{\frac{\ln(1/p)}{2\tau^2}}, \end{aligned}$$

where $\int_{y_0}^\infty e^{-2\tau^2 y^2} dy \leq \int_{y_0}^\infty \frac{y}{y_0} e^{-2\tau^2 y^2} dy = \frac{e^{-2\tau^2 y_0^2}}{4\tau^2 y_0}$. Squaring gives (16). Where this paper uses Theorem L, the extremal decoders have $p \in \{q, q(1-q), q^2\}$ (Proposition 9.11, Remark 10.10), and there it is an equality. The script `thmL.sharpness.py` checks the equality cases exactly on small cubes (all $\{0, 1\}$ -valued f with $n \leq 4$, random nonnegative f of degree at most 2 with $n = 4$), the formulas for ANDs, and (16) on test functions (floating point).

9.8 A third function: g_2

With inputs $(w_1, w_2, w_3, u_1, u_2)$ as before, let

$$\begin{aligned} g_2(w, u) = +1 &\iff (u_1, u_2) \in A''_w, \quad \text{where} \\ A''_{+++} &= \{(+, +)\}, \quad A''_{++-} = A''_{+-+} = \emptyset, \quad A''_{+--} = \{\pm 1\}^2 \setminus \{(+, +)\}, \\ A''_{-++} &= \{(+, +), (-, -)\}, \quad A''_{-+-} = \{(+, -)\}, \quad A''_{--+} = \{(-, +)\}, \quad A''_{---} = \{(+, -), (-, +)\}. \end{aligned} \tag{17}$$

Then $h(w) = \frac{1}{2}|A''_w| - 1$, that is, in the order $+++ , ++- , \dots , ---$,

$$h = \left(-\frac{1}{2}, -1, -1, \frac{1}{2}, 0, -\frac{1}{2}, -\frac{1}{2}, 0\right), \quad \mu = -\frac{3}{8}, \quad \sigma = -1, \quad \text{triv}_{g_2} = \frac{11}{16}.$$

A direct count gives $\hat{g}_2(\{w_j\}) = \hat{h}(\{j\}) = -\frac{1}{8}$ for $j = 1, 2, 3$ and $\hat{g}_2(\{u_i\}) = 0$ for $i = 1, 2$ (over $w = +++ , +- -, -++ , -+- , ---$ the inner sums of u_i over A''_w are $1, -1, 0, 1, -1, 0$ for $i = 1$ and $1, -1, 0, -1, 1, 0$ for $i = 2$). So $J = \{1, 2, 3\}$ is a single class. The function g_2 depends on u_1 and u_2 (as $g_2(-, +, +, u) = u_1 u_2$) and on each w_j , it is not $\pm$ a parity and $\mu \neq 0$, so Theorems B–D apply. Its h belongs to **hard288**, in one of the one-class orbits of size 12. The row laws are

$$\pi_+ = \frac{1+h}{5} = \left(\frac{1}{10}, 0, 0, \frac{3}{10}, \frac{1}{5}, \frac{1}{10}, \frac{1}{10}, \frac{1}{5}\right), \quad \pi_- = \frac{1-h}{11} = \left(\frac{3}{22}, \frac{2}{11}, \frac{2}{11}, \frac{1}{22}, \frac{1}{11}, \frac{3}{22}, \frac{3}{22}, \frac{1}{11}\right);$$

each coordinate w_j equals $+1$ with probability $\beta_+ = \frac{2}{5}$ under π_+ and $\beta_- = \frac{6}{11}$ under π_- , and both laws are invariant under exchanging w_2 and w_3 . With $k, \nu, F, C, X, Y, Z, \Gamma_\nu(F)$ and the basis χ_S (built from $\beta_\pm = \frac{2}{5}, \frac{6}{11}$) as in §9.6:

Proposition 9.11. $\Gamma_\nu(F) \leq 0$ for all $k \geq 0, \nu \in \{-1, 1\}^k$ and $F: \{-1, 1\}^k \rightarrow [-1, 1]$. Consequently $s_{g_2}^L = s_{g_2} = \text{triv}_{g_2} = \frac{11}{16} < s_{g_2}^U$; the same holds for $g_{2,m}(w, u_1, \dots, u_{m-3}) = g_2(w, u_1, u_2 \cdots u_{m-3})$ for every $m \geq 5$, and for every g as in Theorem B whose h lies in the orbit of h_{g_2} under the symmetries of Lemma 11.2.

Proof. The consequences follow from the first statement exactly as in the proof of Proposition 9.10 (again $\sigma h(+, -, -) = -\frac{1}{2} < 0$).

Let $P = \frac{1+F}{2}$, $P_1 = P(X)$, $P_2 = P(Y)$, $P_3 = P(Z)$ and $p = \mathbb{E}P$. Both sides of

$$2h(y) = -p_1 - p_2 - p_3 + 2p_2 p_3 + 2p_1(1 - p_2)(1 - p_3) \quad (p_j = \frac{1+y_j}{2})$$

are multilinear and agree at the vertices: the right-hand side takes the values $-1, -2, -2, 1$ at $p = (1, 1, 1), (1, 1, 0), (1, 0, 1), (1, 0, 0)$ and $0, -1, -1, 0$ at $p = (0, 1, 1), (0, 1, 0), (0, 0, 1), (0, 0, 0)$. Since $P_1(1 - P_2)(1 - P_3) \leq P_1(1 - P_2)$ and X, Y, Z have a common law,

$$2\Gamma_\nu(F) \leq -p - 2\mathbb{E}[P_1 P_2] + 2\mathbb{E}[P_2 P_3].$$

The law of (w_1, w_2) on $(++, +-, -+, --)$ is $(\frac{1}{10}, \frac{3}{10}, \frac{3}{10}, \frac{3}{10})$ under π_+ and $(\frac{7}{22}, \frac{5}{22}, \frac{5}{22}, \frac{5}{22})$ under π_- ; the law of (w_2, w_3) is $(\frac{3}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{2})$ under π_+ and $(\frac{5}{22}, \frac{7}{22}, \frac{7}{22}, \frac{3}{22})$ under π_- . Hence

$$\lambda_+^{12} = \frac{-\frac{1}{5} - \frac{1}{25}}{\frac{24}{25}} = -\frac{1}{4}, \quad \lambda_-^{12} = \frac{\frac{1}{11} - \frac{1}{121}}{\frac{120}{121}} = \frac{1}{12}, \quad \lambda_+^{23} = \frac{\frac{3}{5} - \frac{1}{25}}{\frac{24}{25}} = \frac{7}{12}, \quad \lambda_-^{23} = \frac{-\frac{3}{11} - \frac{1}{121}}{\frac{120}{121}} = -\frac{17}{60},$$

and, writing $\mathbb{E}[P_1 P_2]$ and $\mathbb{E}[P_2 P_3]$ spectrally as in the proof of Lemma 9.3 (the terms with $S = \emptyset$ cancel),

$$2\Gamma_\nu(F) \leq -p + 2 \sum_{S \neq \emptyset} a_S \hat{P}(S)^2, \quad a_S = \lambda_S^{23} - \lambda_S^{12}.$$

If $S = \{i\}$ then $a_S = \frac{7}{12} + \frac{1}{4} = \frac{5}{6}$ when $\nu_i = +1$ and $a_S = -\frac{17}{60} - \frac{1}{12} = -\frac{11}{30}$ when $\nu_i = -1$. If $|S| = 2$, a_S is $\frac{49}{144} - \frac{1}{16} = \frac{5}{18}$, or $-\frac{7}{12} \cdot \frac{17}{60} + \frac{1}{4} \cdot \frac{1}{12} = -\frac{13}{90}$, or $\frac{289}{3600} - \frac{1}{144} = \frac{11}{150}$, according as S contains two, one or no indices with $\nu_i = +1$. If $|S| \geq 3$ then $a_S \leq (\frac{7}{12})^{|S|} + (\frac{1}{4})^{|S|} \leq \frac{185}{864} < \frac{5}{18}$. So $a_S \leq \frac{5}{18}$ for every nonempty S other than the singletons of the $\nu = +1$ rows, and with W_+ as in §9

$$\sum_{S \neq \emptyset} a_S \hat{P}(S)^2 \leq \frac{5}{6} W_+ + \frac{5}{18} (\text{Var} P - W_+) = \frac{5}{9} W_+ + \frac{5}{18} (\mathbb{E} P^2 - p^2).$$

As in §9, $W_+ = W^1(\bar{P})$ for the function $\bar{P} = \mathbb{E}[P \mid x_+] \geq 0$ on the $\frac{2}{5}$ -biased cube, which has mean p and $\mathbb{E} \bar{P}^2 \leq \mathbb{E} P^2$. Theorem L with $q = \frac{2}{5}$ (so $\frac{1-2q}{q} = \frac{1}{2}$) gives $W_+ \leq \frac{2}{5} \mathbb{E} P^2 + \frac{1}{2} p^2$. Hence

$$\sum_{S \neq \emptyset} a_S \hat{P}(S)^2 \leq \frac{2}{9} \mathbb{E} P^2 + \frac{5}{18} p^2 + \frac{5}{18} \mathbb{E} P^2 - \frac{5}{18} p^2 = \frac{1}{2} \mathbb{E} P^2,$$

and $2\Gamma_\nu(F) \leq \mathbb{E}[P^2 - P] \leq 0$. □

Every inequality in this proof is an equality for two inequivalent decoders, both with $\Gamma = 0$: the dictator ($k = 1$, $\nu = +1$, $F(z) = z$; $P_1(1 - P_2)P_3 = 0$ because $\pi_+(+ - +) = 0$) and the AND of two labels ($k = 2$, $\nu = (+1, +1)$, $P = \mathbf{1}\{z = (1, 1)\}$, $p = \frac{4}{25}$; $P_1(1 - P_2)P_3 = 0$ because in every row $w_1 = w_3 = +1$ forces $w_2 = +1$; the only coefficient of P above level one is at $S = \{1, 2\}$, where $a_S = \frac{5}{18}$; and P is an equality case of Theorem L). So the mean-independent bound $W_+ \leq \frac{6}{25}$ of Theorem K cannot replace Theorem L here. All numbers in this subsection and in §9.7, the identity (15) on random $f \geq 0$, and the chain $\Gamma \leq \text{reduction bound} = \text{spectral formula} \leq \frac{1}{2} \mathbb{E}[P^2 - P]$ on random decoders are checked in `d12.check.py`.

10 A maximal-correlation criterion: proof of Theorem F

For random variables L, R on a finite probability space let $\rho(L, R)$ be their maximal correlation, as in §6: the supremum of $\mathbb{E}[\phi(L)\psi(R)]$ over real ϕ, ψ with $\mathbb{E}\phi(L) = \mathbb{E}\psi(R) = 0$ and $\mathbb{E}\phi(L)^2 = \mathbb{E}\psi(R)^2 = 1$, and $\rho(L, R) = 0$ if L or R is a.s. constant.

Lemma 10.1 (Witsenhausen [18]). (a) If L and R take at most two values each, then $\rho(L, R) = |\text{Corr}(L, R)|$ (and 0 if one of them is a.s. constant). (b) If the pairs $(L_1, R_1), \dots, (L_k, R_k)$ are independent, then $\rho((L_1, \dots, L_k), (R_1, \dots, R_k)) = \max_r \rho(L_r, R_r)$. Consequently, for all real functions ϕ, ψ ,

$$|\text{Cov}(\phi(L), \psi(R))| \leq \max_r \rho(L_r, R_r) \sqrt{\text{Var} \phi(L) \text{Var} \psi(R)}, \quad L = (L_r)_r, \quad R = (R_r)_r.$$

Proof. (a) A centred function of a non-constant two-valued variable L is a multiple of $L - \mathbb{E}L$, and similarly for R . (b) “ $\geq$ ” is clear (use functions of one pair). Let $\bar{\rho} = \max_r \rho(L_r, R_r)$, let $\mathcal{T}\psi = \mathbb{E}[\psi(R) \mid L]$ and let $\mathcal{T}_r$ be the corresponding operator for the pair (L_r, R_r) . By independence $\mathcal{T} = \bigotimes_r \mathcal{T}_r$. Each $\mathcal{T}_r$ maps constants to the same constants and centred functions of R_r to centred functions of L_r , and on the latter its norm is $\rho(L_r, R_r)$ (for centred ψ , $\|\mathcal{T}_r \psi\| = \sup \mathbb{E}[\phi(L_r)\psi(R_r)]$ over centred unit ϕ). Let $\psi = \sum_S \psi_S$ be the Efron–Stein decomposition with respect to the (product) law of R . Then $\mathcal{T}\psi_S$ lies in the S -component space of the law of L , these spaces are mutually orthogonal, and $\|\mathcal{T}\psi_S\| \leq \prod_{r \in S} \rho(L_r, R_r) \|\psi_S\| \leq \bar{\rho} \|\psi_S\|$ for $S \neq \emptyset$. So $\|\mathcal{T}\psi\| \leq \bar{\rho} \|\psi\|$ for centred ψ , and for centred ϕ we get $\mathbb{E}[\phi(L)\psi(R)] = \mathbb{E}[\phi(L) \mathcal{T}\psi(L)] \leq \bar{\rho} \|\phi\| \|\psi\|$. The last claim is this inequality applied to $\phi - \mathbb{E}\phi$ and $\psi - \mathbb{E}\psi$. □

Proof of Theorem F. Since $s_g^L \geq \text{triv}_g$ always ($\mathbb{E}_x |\gamma(x)| \geq |\mathbb{E}_x \gamma(x)| = |\mu|$ for mean-zero test functions), by Theorem D it suffices to show $\sigma \Gamma_\nu(F) \geq 0$ for all k, ν and class-wise F . Fix them, let $Y_j = F_{c(j)}(C_{\cdot j})$ and $y_j = \mathbb{E} Y_j$ ($j \in J$). The column $C_{\cdot j}$ has independent entries, and for each row r , $\Pr[C_{rj} = 1]$ is the π_{ν_r} -probability of $w_j = 1$, namely $\frac{1+\nu_r(\mu+\hat{h}(\{j\}))}{2(1+\nu_r\mu)}$. So the law of $C_{\cdot j}$ depends on j only through $\hat{h}(\{j\}) = \hat{g}(\{j\})$, that is, only through the class of j , and since the decoder is also a function of the class, y is constant on classes. As h is multilinear without terms of degree ≥ 3 ,

$$\Gamma_\nu(F) = \mathbb{E} h(Y) = h(y) + \sum_{\{i,j\} \subseteq J} \hat{h}(\{i,j\}) \text{Cov}(Y_i, Y_j).$$

For $i \neq j$ the pairs (C_{ri}, C_{rj}) , $r = 1, \dots, k$, are independent, and (C_{ri}, C_{rj}) has the law of (w_i, w_j) under π_{ν_r} , so its maximal correlation is $|\text{Corr}_{\pi_{\nu_r}}(w_i, w_j)| \leq \rho_{ij}$ by Lemma 10.1(a). By Lemma 10.1(b) and $Y_j^2 \leq 1$, $|\text{Cov}(Y_i, Y_j)| \leq \rho_{ij} \sqrt{\text{Var } Y_i \text{Var } Y_j} \leq \rho_{ij} \sqrt{(1-y_i^2)(1-y_j^2)}$. Hence $\sigma \Gamma_\nu(F) \geq \Phi(y) \geq 0$. $\square$

Corollary 10.2. *Let g be as in Theorem B. If $\text{sgn}(\mu)h(w) < 0$ for some vertex $w \in \{-1, 1\}^J$ that is constant on classes, then $s_g^L > \text{triv}_g$. In particular, if the level-one coefficients $\hat{g}(\{j\})$, $j \in J$, are pairwise distinct, then*

$$s_g^L > \text{triv}_g \iff s_g > \text{triv}_g \iff s_g^U > \text{triv}_g \iff \exists w \in \{-1, 1\}^J : \text{sgn}(\mu)h(w) < 0.$$

Proof. The first claim is Theorem C(iii) with $k = 0$ and $y = w$. If all classes are singletons, every vertex is constant on classes; together with $s_g^L \leq s_g \leq s_g^U$ and Theorem C(i),(ii) this gives the equivalences. $\square$

So in Theorems E and F the obstruction to nontriviality necessarily sits at vertices that are not constant on classes. For $|J| = 2$ with two classes, Corollary 10.2 decides everything; with one class the vertices $(1, -1)$ and $(-1, 1)$ are the only ones that matter, and Theorem F can apply.

10.1 An example with two relevant coordinates

Write the inputs of $g_F: \{-1, 1\}^6 \rightarrow \{-1, 1\}$ as $(w_1, w_2, u_1, u_2, u_3, u_4)$, let $\Sigma = u_1 + u_2 + u_3 + u_4$, and put

$$\begin{aligned} g_F(w, u) &= +1 \iff u \in A_w, \quad \text{where} \\ A_{++} &= \{u_1 = u_2, u_3 = u_4\}, \quad A_{--} = \{\Sigma = 0\}, \\ A_{+-} &= \{\Sigma \in \{4, 2, -2\}\}, \quad A_{-+} = \{\Sigma \in \{2, -2, -4\}\}. \end{aligned} \tag{18}$$

Proposition 10.3. *The function g_F depends on all six coordinates and $\mathbb{E} g_F = -\frac{1}{8}$. Its level-one coefficients are $-\frac{1}{16}$ for w_1, w_2 and 0 for $u_1, \dots, u_4$, so $J = \{1, 2\}$ is a single class, and $h = (-\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, -\frac{1}{4})$ in the order $++, +-, -+, --$. Moreover $s_{g_F}^L = s_{g_F} = \text{triv}_{g_F} = \frac{9}{16} < s_{g_F}^U$.*

Proof. Data. $|A_{++}| = 4$, $|A_{+-}| = |A_{-+}| = 1 + 4 + 4 = 9$ and $|A_{--}| = 6$, so $h(w) = 2|A_w|/16 - 1$ is as stated, $\mu = -\frac{1}{8}$ and $\text{triv}_{g_F} = \frac{9}{16}$. The level-one coefficients of h are $\hat{h}(\{1\}) = \hat{h}(\{2\}) = \frac{1}{4}(-\frac{1}{2} + \frac{1}{8} - \frac{1}{8} + \frac{1}{4}) = -\frac{1}{16}$, and $\hat{h}(\{1, 2\}) = \frac{1}{4}(-\frac{1}{2} - \frac{1}{8} - \frac{1}{8} - \frac{1}{4}) = -\frac{1}{4}$. For the u -coordinates, $\hat{g}_F(\{u_i\}) = 2^{-6} \sum_w \sum_u g_F(w, u) u_i = 2^{-5} \sum_w \sum_{u \in A_w} u_i$; this vanishes because A_{++} and A_{--} are invariant under $u \mapsto -u$ and $A_{+-} = -A_{-+}$. The function depends on w_1 (the point $u = (1, 1, 1, 1)$ lies in A_{++} but not in A_{-+}), on w_2 (the point $-u$ lies in A_{++} but not in A_{+-}) and on every u_i (flipping one u_i changes Σ by ± 2 , so it leaves A_{--}). It is not $\pm$ a parity since $\mu \neq 0$.

$s_{g_F}^U > \text{triv}_{g_F}$. $\sigma = -1$ and $\sigma h(+-) = -\frac{1}{8} < 0$, so Theorem C(ii) applies.

$s_{g_F} = \text{triv}_{g_F}$. Since $|J| = 2$, Theorem F applies once $\Phi \geq 0$ on the class-constant points $y = (t, t)$. We have $\pi_+ = \frac{1+h}{4(1+\mu)} = (\frac{1}{7}, \frac{9}{28}, \frac{9}{28}, \frac{3}{14})$ and $\pi_- = \frac{1-h}{4(1-\mu)} = (\frac{1}{3}, \frac{7}{36}, \frac{7}{36}, \frac{5}{18})$. Under π_+ , $\mathbb{E}w_1 = \mathbb{E}w_2 = -\frac{1}{14}$ and $\mathbb{E}w_1w_2 = -\frac{2}{7}$, so $\text{Corr} = (-\frac{2}{7} - \frac{1}{196})/(1 - \frac{1}{196}) = -\frac{19}{65}$; under π_- , $\mathbb{E}w_j = \frac{1}{18}$ and $\mathbb{E}w_1w_2 = \frac{2}{9}$, so $\text{Corr} = (\frac{2}{9} - \frac{1}{324})/(1 - \frac{1}{324}) = \frac{71}{323}$. Hence $\rho_{12} = \frac{19}{65}$. With $h(t, t) = \mu + 2\hat{h}(\{1\})t + \hat{h}(\{1, 2\})t^2 = -\frac{1}{8} - \frac{t}{8} - \frac{t^2}{4}$,

$$\Phi(t, t) = \frac{1}{8} + \frac{t}{8} + \frac{t^2}{4} - \frac{1}{4} \cdot \frac{19}{65}(1 - t^2) = \frac{27}{520} + \frac{t}{8} + \frac{21}{65}t^2,$$

whose discriminant $\frac{1}{64} - 4 \cdot \frac{21}{65} \cdot \frac{27}{520} = -\frac{13919}{270400}$ is negative. So $\Phi > 0$ on $\mathbb{R}$ (its minimum is $\frac{13919}{349440}$), Theorem F gives $s_{g_F}^L = s_{g_F} = \text{triv}_{g_F}$, and $\sigma\Gamma_\nu(F) \geq \frac{13919}{349440}$ for every decoder. $\square$

All numbers in this proof are checked exactly in `gF_check.py`. Unlike Theorem E, Proposition 10.3 has slack: every decoder has $\sigma\Gamma_\nu(F) \geq \frac{13919}{349440} \approx 0.0398$. The dictator decoder gives $\sigma\Gamma_+ = -(\mu + \|h\|_2^2)/(1 + \mu) = \frac{5}{112} \approx 0.0446$ (here $\|h\|_2^2 = \frac{11}{128} < |\mu|$, whereas the boundary case of §11 is $\|h\|_2^2 = |\mu|$), and coordinate ascent over single decoders with $k \leq 6$ (`j2_ascent.py`) finds nothing smaller.

Proof of Corollary F'. For $m \geq 6$ let $g(w, u_1, \dots, u_{m-2}) = g_F(w, u_1, u_2, u_3, u_4u_5 \cdots u_{m-2})$. The product $t = u_4 \cdots u_{m-2}$ is uniform and independent of (w, u_1, u_2, u_3) , so h, μ and the level-one coefficients of w_1, w_2, u_1, u_2, u_3 are those of g_F . For $i \geq 4$ and $m \geq 7$, $u_i = t \cdot c$ with $c = \prod_{4 \leq l \leq m-2, l \neq i} u_l$ uniform and independent of (w, u_1, u_2, u_3, t) , so $\hat{g}(\{u_i\}) = \mathbb{E}[g_F(w, u_1, u_2, u_3, t) t] \mathbb{E}c = 0$. Flipping u_i flips t , so g depends on every coordinate, and g is not $\pm$ a parity since $\mu \neq 0$. The hypotheses and the conclusion of Theorem F and of Theorem C(ii) depend only on h , so the proof of Proposition 10.3 applies verbatim. For $m \leq 4$ there is no such g with $\mu \neq 0$ by Theorem C(v). For $m = 5$ and $|J| \leq 2$, every g as in Theorem B with $\text{sgn}(\mu)h(w) < 0$ for some w has $s_g > \text{triv}_g$ by Proposition 11.1. $\square$

Two relevant coordinates in one class. The candidates for h can be enumerated. If $|J| = 2$, the two coordinates form one class and $|U| = r \geq 1$, then $h = (x, z, z, y)$ in the order $++, +-, -+, --$, with values in $\{-1 + 2^{1-r}j : j = 0, \dots, 2^r\}$, $x \neq y$ (since $\hat{h}(\{1\}) = \frac{x-y}{4}$), $\mu \neq 0$, $|h(w)| < 1$ for some w (as g depends on u), and $\sum_w h(w)/2^{1-r}$ even. For the last condition, $h(w)/2^{1-r} = |\{u : g(w, u) = 1\}| - 2^{r-1}$, so $\sum_w h(w)/2^{1-r} \equiv |g^{-1}(1)| \pmod{2}$, while $0 = 2^m \hat{g}(\{u_1\}) = 2 \sum_{g(x)=1} u_1 \equiv 2|g^{-1}(1)| \pmod{4}$. We call such h *candidates*; every g as in Theorem B with $|J| = 2$, one class and $|U| = r$ has a candidate h .

Theorem F alone does not decide all candidates (see the table below). For one class of size two the following refinement of Theorem F replaces the bound $\text{Var } Y \leq 1 - y^2$ on the most dangerous part of the covariance by Theorem L.

Proposition 10.4 (refined criterion for one class of size two). *Let g be as in Theorem B with $J = \{1, 2\}$ a single class, $\sigma = \text{sgn}(\mu)$, $b = \hat{h}(\{1, 2\}) \neq 0$ and $\varepsilon = \text{sgn}(\sigma b)$. For $\nu \in \{-1, 1\}$ let $\lambda_\nu = \text{Corr}_{\pi_\nu}(w_1, w_2)$ and $q_\nu = \min_{c=\pm 1} \pi_\nu(w_1 = c)$.*

- (i) *Suppose that $\varepsilon\lambda_{\nu_*} < 0 \leq \varepsilon\lambda_{-\nu_*}$ for some $\nu_* \in \{-1, 1\}$. Put $a = |\lambda_{\nu_*}|$, $\lambda' = |\lambda_{-\nu_*}|$, and $(a_2, M') = (a^3, a\lambda')$ if $\varepsilon = 1$, $(a_2, M') = (a^2, \lambda'^2)$ if $\varepsilon = -1$. Suppose that $a > \max(a_2, M')$, put $q = q_{\nu_*}$ and $\beta = \frac{1-2q}{q}$, and let A be any number with $\max\{a_2, \frac{M'-aq}{1-q}\} \leq A < a$. With*

$$B(p) = \min \{qp + \beta p^2, q(1-p) + \beta(1-p)^2, p(1-p)\},$$

$$\Phi^+(p) = \sigma h(2p-1, 2p-1) - 4|b|[(a-A)B(p) + Ap(1-p)],$$

every k, ν and F satisfy $\sigma\Gamma_\nu(F) \geq \Phi^+(\frac{1+\mathbb{E}F(C_1)}{2})$.

(ii) If $\varepsilon\lambda_\nu \geq 0$ for both ν , put $M = 0$ if $\varepsilon = 1$ and $M = \max(\lambda_+^2, \lambda_-^2)$ if $\varepsilon = -1$; in the situation of (i) with $a \leq \max(a_2, M')$, put $M = \max(a_2, M')$; if $\varepsilon\lambda_\nu < 0$ for both ν , put $M = \max(|\lambda_+|, |\lambda_-|)$. Then the conclusion of (i) holds with $\Phi^+(p) = \sigma h(2p-1, 2p-1) - 4|b|M p(1-p)$.

In both cases, if $\Phi^+ \geq 0$ on $[0, 1]$ then $s_g^L = s_g = \text{triv}_g$.

Proof. By Theorem D it suffices to bound $\sigma\Gamma_\nu(F)$ for a single decoder F (there is one class), and $s_g^L = \text{triv}_g$ once $s_g = \text{triv}_g$, as $\text{triv}_g \leq s_g^L \leq s_g$. Let $X = C_{.1}$ and $Z = C_{.2}$. They have the same law, the product over the rows r of the law of w_1 under π_{ν_r} . Let $F = \sum_S F_S$ be the Efron–Stein decomposition of F with respect to this law. In row r , (X_r, Z_r) is a pair of two-valued variables with equal marginals and correlation λ_{ν_r} , so $\mathbb{E}[\chi(Z_r) | X_r] = \lambda_{\nu_r}\chi(X_r)$ for the centred function χ of one variable, and $\mathbb{E}[\phi(Z_r) | X_r] = \phi$ for constants. As the rows are independent, $\mathbb{E}[F_S(Z) | X] = \lambda_S F_S(X)$ with $\lambda_S = \prod_{r \in S} \lambda_{\nu_r}$, and therefore $\text{Cov}(F(X), F(Z)) = \sum_{S \neq \emptyset} \lambda_S \|F_S\|^2$. With $y = \mathbb{E}F(X)$,

$$\sigma\Gamma_\nu(F) = \sigma h(y, y) + |b| \sum_{S \neq \emptyset} \varepsilon \lambda_S \|F_S\|^2.$$

Call S with $\varepsilon \lambda_S < 0$ dangerous; the other terms are ≥ 0 and are dropped below.

(i) Let R_* and R' be the sets of rows r with $\nu_r = \nu_*$ and $\nu_r = -\nu_*$, and put $S_* = S \cap R_*$, $S' = S \cap R'$. If $\varepsilon = 1$ then $\lambda_S = (-a)^{|S_*|} \lambda'^{|S'|}$, so a dangerous S has $|S_*|$ odd, and $|\lambda_S| = a$ if S is a singleton, $|\lambda_S| \leq a^3 = a_2$ if $S = S_*$ and $|S| \geq 2$ (then $|S| \geq 3$), and $|\lambda_S| \leq a\lambda' = M'$ if $S' \neq \emptyset$. If $\varepsilon = -1$ then $\lambda_S = a^{|S_*|} (-\lambda')^{|S'|}$, so a dangerous S has $|S'|$ even, and $|\lambda_S| = a$ if S is a singleton (then $S = S_*$), $|\lambda_S| \leq a^2 = a_2$ if $S = S_*$ and $|S| \geq 2$, and $|\lambda_S| \leq \lambda'^2 = M'$ if $S' \neq \emptyset$ (then $|S'| \geq 2$). Let $G = \mathbb{E}[F | X_{R_*}] = \sum_{S' \neq \emptyset} F_{S'}$ and let $W_1 = \sum_{r \in R_*} \|F_{\{r\}}\|^2$ be its level-one weight; then $\text{Var } G = \sum_{\emptyset \neq S \subseteq R_*} \|F_S\|^2$ and $\sum_{S' \neq \emptyset} \|F_{S'}\|^2 = \mathbb{E}F^2 - \mathbb{E}G^2$. Hence

$$\sigma\Gamma_\nu(F) \geq \sigma h(y, y) - |b| [aW_1 + a_2(\text{Var } G - W_1) + M'(\mathbb{E}F^2 - \mathbb{E}G^2)],$$

and since $W_1 \leq \text{Var } G$ and $A \geq a_2$, this also holds with A in place of a_2 . Let $P = \frac{1+F}{2}$, $\bar{P} = \frac{1+G}{2} = \mathbb{E}[P | X_{R_*}]$, $p = \mathbb{E}P = \frac{1+y}{2}$, $e = \mathbb{E}P^2 \leq p$ and $f = \mathbb{E}\bar{P}^2 \in [p^2, e]$. Then $W_1 = 4W^1(\bar{P})$, $\text{Var } G = 4(f - p^2)$ and $\mathbb{E}F^2 - \mathbb{E}G^2 = 4(e - f) \leq 4(p - f)$, so

$$\sigma\Gamma_\nu(F) \geq \sigma h(y, y) - 4|b| [(a - A)W^1(\bar{P}) + A(f - p^2) + M'(p - f)].$$

The coordinates X_r , $r \in R_*$, are i.i.d. with $\Pr[X_r = 1] \in \{q, 1 - q\}$; negating them if necessary (which does not change W^1) we are on the q -biased cube. $\bar{P}$ takes values in $[0, 1]$ and has mean p , so Theorem L applied to $\bar{P}$ and to $1 - \bar{P}$ gives $W^1(\bar{P}) \leq qf + \beta p^2$ and $W^1(\bar{P}) \leq q(1 - 2p + f) + \beta(1 - p)^2$; also $W^1(\bar{P}) \leq \text{Var } \bar{P} = f - p^2$. Let $U(f)$ be the smallest of these three affine functions of f ; note $U(p) = B(p)$. As $a - A > 0$, the bracket is at most $V(f) = (a - A)U(f) + A(f - p^2) + M'(p - f)$, a concave piecewise-affine function of f with slopes $(a - A)q + A - M'$ and $(a - A) + A - M' = a - M'$. The first is ≥ 0 because $A(1 - q) \geq M' - aq$, the second is > 0 , so V is nondecreasing and $V(f) \leq V(p) = (a - A)B(p) + Ap(1 - p)$. This gives $\sigma\Gamma_\nu(F) \geq \Phi^+(p)$.

(ii) Every dangerous S has $|\lambda_S| \leq M$: for $\varepsilon = 1$ and $\lambda_\pm \geq 0$ there is none; for $\varepsilon = -1$ and $\lambda_\pm \leq 0$ a dangerous S has even size ≥ 2 , so $|\lambda_S| \leq \max(\lambda_+^2, \lambda_-^2)$; in the situation of (i) this is the case analysis above, with $a \leq M$; and always $|\lambda_S| \leq \max(|\lambda_+|, |\lambda_-|)$. Hence $\sigma\Gamma_\nu(F) \geq \sigma h(y, y) - |b|M \text{Var } F \geq \sigma h(y, y) - 4|b|M p(1 - p)$.

Finally, if $\Phi^+ \geq 0$ on $[0, 1]$, no k, ν, F give $\sigma\Gamma_\nu(F) < 0$, so $s_g = \text{triv}_g$ by Theorem D. $\square$

For $q \leq \frac{1}{2}$ the first, third and second bounds in $B(p)$ are the smallest on $[0, q]$, $[q, 1 - q]$ and $[1 - q, 1]$ respectively (the first is at most $p(1 - p)$ exactly when $p \leq q$), so Φ^+ is a quadratic on each of these intervals and its nonnegativity is an exact finite check. Since $B(p) \leq p(1 - p)$, $A < a \leq \rho_{12}$ and, in (ii), $M \leq \rho_{12}$, we have $\Phi^+(p) \geq \Phi(2p - 1, 2p - 1)$: Proposition 10.4 is never weaker than Theorem F. In (i), $\Phi^+(p) - \Phi(2p - 1, 2p - 1) = 4|b|[(\rho_{12} - a)p(1 - p) + (a - A)(p(1 - p) - B(p))]$, so it gains at least where $B(p) < p(1 - p)$, that is, for $p < q$ and $p > 1 - q$. The smallest admissible A gives the strongest bound. At $p = q$ the bound of (i) is attained by a dictator decoder ($k = 1$, $\nu = \nu_*$, $F(z) = \pm z$ with $\mathbb{E}F = 2q - 1$), whose value is $\sigma h(2q - 1, 2q - 1) - 4|b|a q(1 - q)$.

For $p < q$ Proposition 10.4 loses in two places: it discards the level-two weight inside the rows of sign ν_* , which (for $\varepsilon = 1$) has the favourable coefficient a^2 , and it bounds the sets that meet the other rows separately, with coefficient M' . Both losses disappear if the column pair is compared directly with a second coupling of the same columns. This is the argument of Theorem L, applied to P itself, with the anti-coupling used only in the rows of sign ν_* .

Lemma 10.5 (coupling bound). *Let g be as in Theorem B with $J = \{1, 2\}$ a single class, $\sigma = \text{sgn}(\mu)$, $b = \hat{h}(\{1, 2\}) \neq 0$ and $\varepsilon = \text{sgn}(\sigma b)$. Let λ_ν and q_ν be as in Proposition 10.4, $r_\nu = \frac{q_\nu}{1 - q_\nu}$, let $t_\nu \in [-r_\nu, 1]$ for $\nu \in \{-1, 1\}$, and let $c, \kappa \geq 0$ satisfy*

$$-\varepsilon \lambda_1^i \lambda_{-1}^j \leq \kappa - c t_1^i t_{-1}^j \quad \text{for all integers } i, j \geq 0 \text{ with } (i, j) \neq (0, 0). \quad (19)$$

Then every k, ν and F satisfy, with $p = \frac{1 + \mathbb{E}F(C_1)}{2}$,

$$\sigma \Gamma_\nu(F) \geq \sigma h(2p - 1, 2p - 1) - 4|b|[\kappa p(1 - p) + c \min(p, 1 - p)^2].$$

Proof. Let $X = C_1$, $Z = C_2$, $P = \frac{1 + F}{2}$, and let $P = \sum_S P_S$ be the Efron–Stein decomposition with respect to the law of X . As in the proof of Proposition 10.4, $\mathbb{E}[P(X)P(Z)] = \sum_S \lambda_S \|P_S\|^2$ with $\lambda_S = \prod_{r \in S} \lambda_{\nu_r}$, so

$$\sigma \Gamma_\nu(F) = \sigma h(2p - 1, 2p - 1) + 4|b| \varepsilon \sum_{S \neq \emptyset} \lambda_S \|P_S\|^2.$$

A row r in which w_1 is π_{ν_r} -almost surely constant lies in no S with $P_S \neq 0$, so we may assume $q_{\nu_r} > 0$ for all r . In row r , the joint laws of two $\{-1, 1\}$ -valued variables that both have the law of w_1 under π_{ν_r} form a segment: it runs from the law under which the two variables never both take the less likely value (correlation $-r_{\nu_r}$) to the diagonal law (correlation 1), and the correlation is affine along it. So there is a random point $\tilde{Z}$ such that the pairs $(X_r, \tilde{Z}_r)$ are independent over r , $\tilde{Z}_r$ has the law of X_r , and $\text{Corr}(X_r, \tilde{Z}_r) = t_{\nu_r}$. The computation for Z gives $\mathbb{E}[P(X)P(\tilde{Z})] = \sum_S t_S \|P_S\|^2$ with $t_S = \prod_{r \in S} t_{\nu_r}$, and $\mathbb{E}[P(X)P(\tilde{Z})] \geq 0$ because $P \geq 0$. If $S \neq \emptyset$ has i rows of sign 1 and j rows of sign -1 , then (19) says $-\varepsilon \lambda_S \leq \kappa - c t_S$. Hence

$$-\varepsilon \sum_{S \neq \emptyset} \lambda_S \|P_S\|^2 \leq \kappa \sum_{S \neq \emptyset} \|P_S\|^2 - c \sum_{S \neq \emptyset} t_S \|P_S\|^2 \leq \kappa \text{Var } P + c p^2 \leq \kappa p(1 - p) + c p^2,$$

since $\sum_{S \neq \emptyset} t_S \|P_S\|^2 = \mathbb{E}[P(X)P(\tilde{Z})] - p^2 \geq -p^2$ and $\text{Var } P \leq \mathbb{E}P - p^2$. The same argument applied to $1 - P$, whose components for $S \neq \emptyset$ are $-P_S$ and whose mean is $1 - p$, gives the bound with $(1 - p)^2$ in place of p^2 . $\square$

Proposition 10.6 (coupling criterion). *In the setting of Lemma 10.5 suppose that $\varepsilon = 1$ and $\lambda_{\nu_*} < 0 \leq \lambda_{-\nu_*}$ for some $\nu_* \in \{-1, 1\}$. Put $a = -\lambda_{\nu_*}$, $q = q_{\nu_*}$, $r = \frac{q}{1-q}$ and*

$$\theta = \sup_{\ell \geq 1 \text{ odd}} \frac{r^\ell - a^\ell}{1 + r^\ell},$$

$$\Phi^\sharp(p) = \sigma h(2p - 1, 2p - 1) - 4|b| a \min \left\{ p(1 - p), \theta p(1 - p) + \frac{1 - \theta}{r} \min(p, 1 - p)^2 \right\}.$$

Then $0 \leq \theta \leq q$, and every k , ν and F satisfy $\sigma \Gamma_\nu(F) \geq \Phi^\sharp(\frac{1 + \mathbb{E}F(C.1)}{2})$. If $\Phi^\sharp \geq 0$ on $[0, 1]$, then $s_g^L = s_g = \text{triv}_g$.

Proof. Two $\{-1, 1\}$ -valued variables with the same law, whose less likely value has probability q , have correlation at least $-r$; so $a \leq r \leq 1$. Hence $0 \leq \frac{r^\ell - a^\ell}{1 + r^\ell} \leq \frac{r^\ell}{1 + r^\ell} \leq \frac{r}{1 + r} = q$ for every odd ℓ , so $0 \leq \theta \leq q$. (If $r < 1$ the terms tend to 0 and the supremum is a maximum; if $r = 1 > a$ then $\theta = \frac{1}{2} = q$. The case $r = a = 1$ does not occur: $a = 1$ means $w_1 = -w_2$ almost surely under π_{ν_*} , so $\pi_{\nu_*}(++) = \pi_{\nu_*}(--) = 0$ and $h(++) = h(--) = -\nu_*$, whereas $h(++) - h(--) = 4\hat{h}(\{1\}) \neq 0$ because $h(+-) = h(-+)$ for one class.) Let $\theta' \in [\theta, 1]$ and apply Lemma 10.5 with $t_{\nu_*} = -r$, $t_{-\nu_*} = \lambda' := \lambda_{-\nu_*} \in [0, 1]$, $c = \frac{a(1 - \theta')}{r}$ and $\kappa = a\theta'$. For S with i rows of sign ν_* and j rows of sign $-\nu_*$, condition (19) (with the roles of i and j exchanged if $\nu_* = -1$) reads

$$D_{ij} := a\theta' + \lambda'^j ((-a)^i - a(1 - \theta')(-1)^i r^{i-1}) \geq 0.$$

For $i = 0$, $D_{0j} = a\theta' + \lambda'^j (1 - \frac{a(1 - \theta')}{r}) \geq 0$ as $a \leq r$. For $i \geq 1$ put $B_i = a^{i-1} - (1 - \theta')r^{i-1}$, so that $D_{ij} = a\theta' - (-1)^{i-1} \lambda'^j a B_i$. For odd i , $B_i \leq r^{i-1} - (1 - \theta')r^{i-1} \leq \theta'$, so $D_{ij} \geq a\theta'(1 - \lambda'^j) \geq 0$ if $B_i > 0$, and $D_{ij} \geq a\theta'$ otherwise. For even i , $D_{ij} \geq a \min\{\theta', \theta' + B_i\}$ since $0 \leq \lambda' \leq 1$, and $\theta' + B_i \geq 0$ because $\theta' \geq \theta \geq \frac{r^{i-1} - a^{i-1}}{1 + r^{i-1}}$ ($i - 1$ is odd). So the lemma applies and gives $\sigma \Gamma_\nu(F) \geq \sigma h(2p - 1, 2p - 1) - 4|b| a [\theta' p(1 - p) + \frac{1 - \theta'}{r} \min(p, 1 - p)^2]$. The bracket is affine in θ' ; $\theta' = \theta$ and $\theta' = 1$ give the two terms of the minimum. The last sentence follows from Theorem D, as in Proposition 10.4. $\square$

For $p \leq \frac{1}{2}$ the second term of the minimum is the smaller one exactly when $p \leq q$ (both terms agree at $p = q$), so $\Phi^\sharp$ is a quadratic on each of $[0, q]$, $[q, 1 - q]$ and $[1 - q, 1]$, and $\Phi^\sharp(q)$ is the value of the dictator decoder. Proposition 10.6 is never weaker than Proposition 10.4(i) when $\varepsilon = 1$. On $[q, 1 - q]$ both brackets equal $a p(1 - p)$. On $[0, q]$, $\theta' = q$ is admissible and gives the bracket $a(qp(1 - p) + \frac{(1 - q)^2}{q} p^2) = a(qp + \beta p^2)$, which is at most $(a - A)(qp + \beta p^2) + A p(1 - p)$ because $qp + \beta p^2 \leq p(1 - p)$ there, with strict inequality for $0 < p < q$ (and $A \geq a_2 > 0$); $[1 - q, 1]$ is symmetric. The choice $\theta' = \theta$ gains more, as $\theta < q$ whenever $r < 1$ (equality in $\frac{r^\ell}{1 + r^\ell} \leq q$ needs $\ell = 1$, and then $a > 0$ makes the term smaller); and no separate coefficient M' is needed for the rows of sign $-\nu_*$. The bracket of Proposition 10.6 is the best bound obtainable from the inequalities used in its proof (Remark 10.10 at the end of this section).

Proposition 10.7 (computer-assisted, exact). *Every candidate has $\mu \neq 0$, and $-h$ is a candidate whenever h is. The candidates with $\mu < 0$ fall into the following classes: TRIV ($\sigma h \geq 0$, Theorem C(i)), CERT (a certificate of Theorem C(iii) with a constant decoder, a $[-1, 1]$ -valued decoder with $k = 1$, or a $\{-1, 1\}$ -valued decoder with $k = 2$), F (Theorem F applies), F^+ (Theorem F does not apply, and Proposition 10.4 applies), $F^\sharp$ (Theorem F and Proposition 10.4 do not apply, and Proposition 10.6 applies), CERT' (none of the above, and a certificate with a $[-1, 1]$ -valued decoder with $k = 2$ or $k = 3$ exists) and UNDEC (none of these). For h in F , F^+ or $F^\sharp$, every g with this h has $s_g^L = s_g = \text{triv}_g < s_g^U$; for h in CERT or CERT', every g with this h has $s_g > \text{triv}_g$.*

	TRIV	CERT	F	F ⁺	F [#]	CERT'	UNDEC
$r = 3$ ($m = 5$)	40	86	0	0	0	0	0
$r = 4$ ($m = 6$)	288	718	14	2	0	0	0
$r = 5$ ($m = 7$)	2176	5818	186	10	0	0	0
$r = 6$ ($m = 8$)	16896	46696	1806	114	18	4	0

Proof. `maxcorr_J2.py`, in exact rational arithmetic. For one class the two columns have the same law, so ρ_{12} and $\lambda_{\pm}$ are rational, and $\Phi(t, t)$ and the three pieces of Φ^+ and of $\Phi^{\#}$ are quadratics with rational coefficients, minimised exactly on their intervals. For F^+ the script tests Proposition 10.4(i) with $A = \max\{a_2, \frac{M'-aq}{1-q}\}$, which suffices because $B(p) \leq p(1-p)$ makes Φ^+ nonincreasing in A , and the first two cases of (ii). It does not test the last case of (ii) (two dangerous signs), and this does not affect the table: such a candidate would fail the test for $F^{\#}$, whose hypotheses include exactly one dangerous sign, and so land in $CERT'$ or $UNDEC$; but Proposition 10.4 cannot apply to a candidate in $CERT'$ (by Theorem D), and $UNDEC$ is empty for $r \leq 6$. The number θ of Proposition 10.6 is rational: $\theta = 0$ if $a = r$, $\theta = \frac{1}{2}$ if $r = 1 > a$, and otherwise the first term is positive and every term with exponent ℓ is below r^{ℓ} , so only the finitely many terms with r^{ℓ} above the first term can matter. The $k = 1$ values form a quadratic in $(F(1), F(-1))$, maximised exactly over the square; the 48 values of $\{-1, 1\}$ -valued decoders with $k = 2$ (16 decoders for each of the row-sign patterns $(+, +)$, $(+, -)$ and $(-, -)$) are screened in floating point and every decoder whose screened value exceeds -10^{-9} is re-evaluated exactly. For $CERT'$ the maximum of Γ_{ν} over $[-1, 1]$ -valued decoders with $k \leq 3$ is computed exactly by the face enumeration described in the proof of Proposition 10.9; the four candidates are $(-\frac{9}{32}, \frac{1}{8}, \frac{1}{8}, -\frac{3}{32})$ and $(-\frac{5}{32}, \frac{1}{16}, \frac{1}{16}, -\frac{1}{32})$, discussed below, and their images under $w \mapsto -w$, and their certificates are re-checked in `m8_check.py`. For h in F , F^+ or $F^{\#}$, $\sigma h(w) < 0$ for some w (otherwise h would be $TRIV$), so $s_g^U > \text{triv}_g$ by Theorem C(ii). For $CERT$ and $CERT'$, $s_g > \text{triv}_g$ by Theorem C(iii). $\square$

Every candidate in F^+ also satisfies the hypotheses of Proposition 10.6 (in particular $\varepsilon = 1$), with $\Phi^{\#} \geq 0$, as the comparison above predicts (`maxcorr_J2.py` runs both tests on every candidate outside $TRIV$, $CERT$ and F); all candidates in $F^{\#}$ and $CERT'$ have $\varepsilon = 1$ and exactly one dangerous sign.

Corollary 10.8 (computer-assisted). *Let g be as in Theorem B with $|J| \leq 2$ and $m \leq 8$. Then $s_g > \text{triv}_g$ if and only if Theorem D(c) holds with some $k \leq 3$. If $m \leq 7$, then $k \leq 2$ suffices; for $m = 8$ it does not (Proposition 10.9).*

Proof. If $\sigma h \geq 0$ on $\{-1, 1\}^J$, then $s_g = \text{triv}_g$ by Theorem C(i). Otherwise $\sigma h(w) < 0$ for some w . If $|J| \leq 1$, or $|J| = 2$ and $\hat{g}(\{1\}) \neq \hat{g}(\{2\})$, then w is constant on classes and Theorem C(iii) holds with $k = 0$. Let $|J| = 2$ be one class. For $m \leq 4$ the proof of Theorem C(v) gives a certificate with $k \leq 1$. For $5 \leq m \leq 8$, $|U| = r = m - 2 \in \{3, 4, 5, 6\}$ and h is a candidate; replacing g by $-g$ if necessary (this replaces h , σ and π_{ν} by $-h$, $-\sigma$ and $\pi_{-\nu}$, hence λ_{ν} , q_{ν} by $\lambda_{-\nu}$, $q_{-\nu}$ and b by $-b$, and leaves σh , ε , the set of values $\sigma \Gamma_{\nu}(F)$ and the hypotheses of Theorem F and Propositions 10.4 and 10.6 unchanged) we may take $\mu < 0$. By Proposition 10.7, h is in $CERT$, which is a certificate with $k \leq 2$, in $CERT'$, which is a certificate with $k \leq 3$ and occurs only for $m = 8$, or in F , F^+ or $F^{\#}$, in which case $s_g = \text{triv}_g$ and, by Theorem D, there is no certificate at all. $\square$

For $m = 9$ the same methods leave 34 candidates open; see §11.

Examples. The two F^+ candidates for $m = 6$ are $h = (-\frac{5}{8}, \frac{1}{8}, \frac{1}{8}, -\frac{1}{8})$ and its image $(-\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, -\frac{5}{8})$ under $w \mapsto -w$. For the first, $\min_t \Phi(t, t) = -\frac{1}{1344}$, attained at $t = -\frac{8}{21}$, so Theorem F does not apply; coordinate ascent over single decoders with $k \leq 6$ (`j2_ascent.py`) finds no value of Γ

above $-\frac{1}{56}$, the value of the dictator decoder. In Proposition 10.4(i), $\sigma = -1$, $b = -\frac{1}{4}$, $\varepsilon = 1$, $\lambda_+ = -\frac{5}{16}$ and $\lambda_- = \frac{17}{80}$, so $\nu_* = +1$, $a = \frac{5}{16}$, $\lambda' = \frac{17}{80}$, $a_2 = a^3$, $M' = a\lambda' = \frac{17}{256}$, $q = \frac{3}{7}$, $\beta = \frac{1}{3}$, and $\frac{M'-aq}{1-q} < 0$. Already $A = M'$ works: $256\Phi^+(p) = 32 - 172p + 252p^2$ on $[0, \frac{3}{7}]$ (negative discriminant), $16\Phi^+(p) = 2 - 13p + 21p^2$ on $[\frac{3}{7}, \frac{4}{7}]$ (roots $\frac{2}{7}$ and $\frac{1}{3}$), and $64\Phi^+(p) = -4 - 19p + 63p^2$ on $[\frac{4}{7}, 1]$, which is increasing there and equals $\frac{40}{7}$ at $p = \frac{4}{7}$. The minimum of Φ^+ is $\frac{167}{16128}$, at $p = \frac{43}{126}$; at the dictator value $p = \frac{3}{7}$ it is $\frac{1}{56}$. The most delicate of the ten F^+ candidates for $m = 7$ is $h = (-\frac{3}{4}, \frac{1}{8}, \frac{1}{8}, -\frac{1}{8})$, with its mirror image, where the freedom in A is needed. Here $\sigma = -1$, $b = -\frac{9}{32}$, $\varepsilon = 1$, $\lambda_+ = -\frac{67}{176}$, $\lambda_- = \frac{11}{48}$, $q = \frac{11}{27}$, $\beta = \frac{5}{11}$, $M' = \frac{67}{768}$ and $\frac{M'-aq}{1-q} = -\frac{469}{4096}$, so $A = a^3 = (\frac{67}{176})^3$ is admissible. The minimum of Φ^+ is $\frac{32815487183}{30334366605312} \approx 1.08 \cdot 10^{-3}$ on $[0, \frac{11}{27}]$, $\frac{1}{216}$ on $[\frac{11}{27}, \frac{16}{27}]$ (at the dictator value $p = \frac{11}{27}$) and $\frac{13}{108}$ on $[\frac{16}{27}, 1]$. With $A = M'$ the minimum on $[0, \frac{11}{27}]$ would be $-\frac{2161}{43079040} < 0$, and Theorem F gives $\min \Phi \approx -0.0137$. All these numbers are checked exactly in `fplus_check.py`.

Eight variables. For $r = 6$, Theorem F and Proposition 10.4 leave 22 candidates, forming 11 pairs $\{h, h(-\cdot)\}$. For each of them $\varepsilon = 1$, exactly one sign is dangerous, and Φ^+ is negative only on $[0, q]$ (only on $[1 - q, 1]$ for the mirror images). Proposition 10.6 settles 9 of the pairs. The tightest is $h = (-\frac{23}{32}, \frac{1}{8}, \frac{1}{8}, -\frac{3}{32})$: here $\sigma = -1$, $b = -\frac{17}{64}$, $\lambda_+ = -\frac{23}{65}$, $\lambda_- = \frac{163}{747}$, so $\nu_* = +1$, $a = \frac{23}{65}$, $q = \frac{9}{22}$, $\frac{q}{1-q} = \frac{9}{13}$, and $\theta = \frac{3589}{16625}$ is the term $\ell = 3$ (the term $\ell = 1$ is $\frac{1}{5}$). With $\sigma h(2p - 1, 2p - 1) = \frac{3}{32} - \frac{7}{16}p + \frac{17}{16}p^2$, on $[0, \frac{9}{22}]$

$$\Phi^\sharp(p) = \frac{3}{32} - \frac{4483837}{8645000}p + \frac{6981458}{9725625}p^2,$$

with minimum $\frac{120965046879}{1931350541120000} \approx 6.3 \cdot 10^{-5}$ at $p = \frac{40354533}{111703328}$; on $[q, 1 - q]$ the minimum is $\frac{3}{1760}$, the dictator value at $p = q$, and on $[1 - q, 1]$ it is larger (`fsharp_check.py`, which also lists the minima of $\Phi^\sharp$ for all 18 candidates). The other two pairs are not trivial. For $h = (-\frac{5}{32}, \frac{1}{16}, \frac{1}{16}, -\frac{1}{32})$ the decoder with $k = 2$, $\nu = (+, +)$ and $F(++ , +-, -+, --) = (-\frac{554}{1215}, -1, 1, -1)$ has $\Gamma_\nu(F) = \frac{67561}{411505920} > 0$, although no $\{-1, 1\}$ -valued decoder with $k \leq 2$ is a certificate. For $h = (-\frac{9}{32}, \frac{1}{8}, \frac{1}{8}, -\frac{3}{32})$ every decoder with $k \leq 2$ has $\Gamma < 0$, and a certificate needs $k = 3$:

Proposition 10.9 (computer-assisted, exact). *Let $v = (1, \dots, 1) \in \{-1, 1\}^6$, let $u^{(1)}, \dots, u^{(31)}$ be the points of $\{-1, 1\}^6$ with first coordinate 1 other than v , in lexicographic order with 1 before -1 , and put $Q_n = \{\pm u^{(1)}, \dots, \pm u^{(n)}\}$. Define $g_3: \{-1, 1\}^2 \times \{-1, 1\}^6 \rightarrow \{-1, 1\}$ by $g_3(w, u) = 1$ iff $u \in A_w$, where*

$$A_{++} = \{v\} \cup Q_{11}, \quad A_{+-} = A_{-+} = Q_{18}, \quad A_{--} = \{-v\} \cup Q_{14}.$$

Then g_3 depends on all 8 coordinates, $\mathbb{E}g_3 = -\frac{1}{32}$, $\hat{g}_3(\{1\}) = \hat{g}_3(\{2\}) = -\frac{3}{64}$, $\hat{g}_3(\{j\}) = 0$ for $j \geq 3$, and $h = (-\frac{9}{32}, \frac{1}{8}, \frac{1}{8}, -\frac{3}{32})$. Every decoder with $k \leq 2$ has $\Gamma_\nu(F) \leq -\frac{181991}{5205698560} < 0$, with equality for $k = 2$, $\nu = (+, +)$, $F = (\frac{2517}{5290}, -1, 1, -1)$ and its label swap, while the decoder with $k = 3$, $\nu = (+, +, +)$ and

$$F(z) = \begin{cases} -\frac{41937}{121670}, & z = (+, +, +), \\ 1, & z_3 = +1, (z_1, z_2) \neq (+, +), \\ -1, & z_3 = -1 \end{cases}$$

has $\Gamma_\nu(F) = \frac{4317675409}{14846652293120} > 0$. Hence $\text{triv}_{g_3} = \frac{33}{64} < s_{g_3} \leq s_{g_3}^U$, and Theorem D(c) holds for g_3 with $k = 3$ but not with $k \leq 2$.

Proof. `m8_check.py`, in exact rational arithmetic: it builds g_3 , checks the Fourier coefficients and the dependence on every coordinate, computes h and the laws $\pi_\pm$ from g_3 , and evaluates Γ for

the decoder above. For $k \leq 2$, $\Gamma_\nu(F)$ is a quadratic in the 2^k values of F . If its maximum over $[-1, 1]^{2^k}$ is attained at a point x in the relative interior of a face, then x is a critical point of the restriction to that face; if the Hessian of the restriction is singular, the quadratic is constant on a line through x inside the face, so the maximum is also attained on a smaller face. So the maximum is the largest value at the critical points of the faces with nonsingular restricted Hessian, which are rational and are found by enumerating the 3^{2^k} faces (`j2_k2_exact.py`); for each ν with $k \leq 2$ the maximum is negative. As $\sigma = -1$, $\Gamma > 0$ is a certificate, so $s_{g_3} > \text{triv}_{g_3}$ by Theorem C(iii); for one class, class-wise decoders are single decoders. $\square$

The two certificates have the same shape. With $c = +1$, the less likely value of w_1 under π_+ , $P = \frac{1+F}{2}$ equals the dictator $\mathbf{1}[z_k = c]$ except on the event that all labels equal c , where it is smaller. This moves p below q , towards the minimum of $\sigma h(2p - 1, 2p - 1)$, which lies at $p = \frac{7}{20}$ for g_3 and at $p = \frac{3}{10}$ for $h = (-\frac{5}{32}, \frac{1}{16}, \frac{1}{16}, -\frac{1}{32})$, while the dictator itself is not a certificate. Proposition 10.6 is exact for the dictator (at $p = q$), and both certificates have $p < q$.

Remark 10.10 (the bracket of Proposition 10.6 is optimal for the inequalities used). Suppose $r < 1$, let ℓ be an odd index with $\theta = \frac{r^\ell - a^\ell}{1 + r^\ell}$, and let $0 < p \leq q$. Let R_* be the set of rows of sign ν_* , consider decoders with $P_S = 0$ for every $S \not\subseteq R_*$, and let $W_i = \sum_{S \subseteq R_*, |S|=i} \|P_S\|^2$. For them the proof of Proposition 10.6 uses only $W_i \geq 0$, $\sum_i W_i \leq p(1 - p)$ and $\sum_i (-r)^i W_i \geq -p^2$ (the variant with $(1 - p)^2$ is weaker, as $p \leq \frac{1}{2}$), and bounds $\sum_i (-1)^{i+1} a^i W_i = -\sum_{S \neq \emptyset} \lambda_S \|P_S\|^2$ by the bracket $a(\theta p(1 - p) + \frac{1-\theta}{r} p^2)$. This is the maximum under these constraints:

$$W_1 = \frac{p^2 + r^{\ell+1} p(1 - p)}{r(1 + r^\ell)}, \quad W_{\ell+1} = \frac{r p(1 - p) - p^2}{r(1 + r^\ell)}, \quad W_i = 0 \text{ otherwise}$$

satisfies them ($W_{\ell+1} \geq 0$ because $\frac{p}{1-p} \leq r$), with equality in the last two, and $aW_1 - a^{\ell+1}W_{\ell+1} = \frac{a}{1+r^\ell}((r^\ell - a^\ell)p(1 - p) + \frac{1+a^\ell}{r} p^2)$, which is the bracket since $\frac{1-\theta}{r} = \frac{1+a^\ell}{r(1+r^\ell)}$. For $\ell = 1$ these are, at $p = q^2$ and at $p = q(1 - q)$, the level weights $(W_1, W_2) = (2q^3(1 - q), q^2(1 - q)^2)$ of $P = \mathbf{1}[z_1 = z_2 = c]$ and $(W_1, W_2) = (q(1 - q)(1 - 2q + 2q^2), q^2(1 - q)^2)$ of $P = \mathbf{1}[z_1 = \bar{c}, z_2 = c]$, with both rows of sign ν_* , c the less likely value and $\bar{c} = -c$ (two of the equality cases of Theorem L). So if θ is attained at $\ell = 1$, then the bound $\sigma \Gamma_\nu(F) \geq \Phi^\sharp(p)$ of Proposition 10.6 holds with equality for these two decoders, at $p = q^2$ and $p = q(1 - q)$, and for the dictator decoder, at $p = q$.

Theorem F and the boundary functions. Among the 288 functions of §11 only the 48 in the orbits of the representatives called rep3 and rep8 below have $\hat{h}(\{1, 2, 3\}) = 0$ (`class_flips.py`). For both representatives there is a class-constant y with $\Phi(y) < -0.058$ (exact, `thmF_boundary.py`). The minimum of Φ over class-constant points is invariant under the symmetries of Lemma 11.2 ($h \mapsto -h$ exchanges π_+ and π_- , and $h \mapsto h(-\cdot)$ replaces $\Phi(y)$ by $\Phi(-y)$). So Theorem F settles none of the 288 (the orbits rep3 and rep8 are settled in §11 by a re-coupling argument, Proposition 11.5). This is expected: the dictator decoder has $\Gamma = 0$ for all of them, so a bound $\sigma \Gamma \geq \Phi \geq 0$ would force Φ to vanish at the interior class-constant point $y = \mathbb{E}_{\pi_+} w$, together with its derivatives along class-constant directions.

11 What remains open; numerics

Five variables. By Theorems C and D, a g with $\text{sgn}(\mu)h(w) < 0$ for some w has $s_g > \text{triv}_g$ iff some labelled decoder certificate exists, and by (iv) every g without a known certificate has

$\|h\|^2 \leq |\mu|$. Theorem E shows that for $m = 5$ the answer is not always “yes”; the following proposition and Conjecture 11.14 describe what is left.

Proposition 11.1 (computer-assisted). *Let $m = 5$ and let g be as in Theorem B. Then $s_g > \text{triv}_g \iff \exists w : \text{sgn}(\mu)h(w) < 0$, except possibly when $|J| = 3$ and h is one of the 288 functions listed in `numerics/outputs/hard288.txt` (a list closed under $h \mapsto -h$); all of these satisfy $\|h\|_2^2 = |\mu|$.*

Proof. By Theorem C(i),(iii), g is decided once a TRIV certificate or a labelled certificate exists. If $|J| = 0$ then $h \equiv \mu$, so $\text{sgn}(\mu)h \geq 0$ and (i) applies. If $|J| = 5$ then $\|h\|^2 = 1 > |\mu|$ and (iv) applies. If $|J| = 1$, every $w \in \{-1, 1\}^J$ is class-constant, so the criterion itself is an L0 certificate. For $|J| = 4$ there is one u -coordinate, and the claim is the case $|U| = 1$ of the remark after Theorem C. As a cross-check: h takes values in $\{-1, 0, 1\}$, and it is realisable with $\hat{g}(\{u\}) = 0$ and g depending on u iff the number of zeros of h is even and positive. All 3^{16} such h with $\mu \neq 0$ and nonzero level one were enumerated exactly (`q6_d4r1.py`): 5,242,688 admissible h , of which 8192 are TRIV, 4,824,356 have an L0 and 410,140 a LAB1 certificate, and none is undecided. For $|J| = 2$ and $|J| = 3$, `q6_hsearch.py` enumerates all h with values in $\{-1 + 2a/2^r\}$ that satisfy necessary conditions for realisability. It uses floating point, but every quantity it computes is a dyadic rational with denominator dividing $2^{2r+|J|} \leq 2^8$ and modulus at most 2, so every operation is exact and its tolerance 10^{-12} only separates 0 from numbers of modulus at least 2^{-8} . For $|J| = 2$ every such h has a TRIV/L0/LAB1 certificate. For $|J| = 3$ exactly 348 do not; 60 of these have a $k = 2$ certificate (`q6_exact_k2.py`, exact integers), leaving the 288 listed functions. $\square$

The boundary case. An exhaustive search over $h: \{-1, 1\}^3 \rightarrow \{-1, -\frac{1}{2}, 0, \frac{1}{2}, 1\}$ (the possible h with $m = 5$, $|J| = 3$) finds 348 such h not decided by TRIV/L0/LAB1. Each of them is realised by some g on 5 variables with $J = \{1, 2, 3\}$ depending on all coordinates, and each has $\|h\|_2^2 = |\mu|$ exactly (exact check, `q6_realise_check.py`). Exactly 60 of them are decided by $k = 2$ decoders with values in $\{-1, 1\}$; allowing the value 0 as well decides no further h (an empirical outcome of the same computation, not a general fact). For the remaining 288, every class-wise decoder with $k \leq 2$ and values in $\{-1, 0, 1\}$ gives $\sigma\Gamma_\nu(F) \geq 0$ (exhaustive, exact integers, `q6_exact_k2.py`); since Γ_ν is not multilinear in the values $F(z)$, this does not cover all $[-1, 1]$ -valued decoders with $k \leq 2$ (for the 240 functions outside the orbits settled by Theorem E and Propositions 9.10, 9.11 this is done, for $k \leq 3$, in Proposition 11.4 below). Numerical local optimisation over continuous decoders with $k = 3$ and $k = 4$ from random starts also finds no certificate (the best value found is 0).

Lemma 11.2 (symmetries of the criterion). *Whether condition (d) of Theorem D holds for h is unchanged if the coordinates of J are permuted, if h is replaced by $-h$, or if h is replaced by $\tilde{h}(w) = h(-w)$.*

Proof. Permutations act on the columns of C . Replacing h by $-h$ changes the signs of μ and σ and exchanges π_+ and π_- , so $\sigma\Gamma_\nu(F)$ computed for $-h$ equals $\sigma\Gamma_{-\nu}(F)$ computed for h . For $\tilde{h}$, σ is unchanged and $\tilde{\pi}_b(w) = \pi_b(-w)$, so a matrix $\tilde{C}$ with rows $\sim \tilde{\pi}_{\nu_i}$ has the law of $-C$, and $\tilde{h}(F(\tilde{C}_{\cdot j}))_j$ has the law of $h(-F(-C_{\cdot j}))_j = h(\tilde{F}(C_{\cdot j}))_j$ with $\tilde{F}(z) = -F(-z)$ (the multilinear extension of h is $y \mapsto h(-y)$). $\square$

Under the group generated by these operations the 288 functions form 14 orbits, ten with 24 elements and four with 12 (`hard288_orbits.py`); in five orbits (the four of size 12, including that of g_0 , and one of size 24) all level-one coefficients of h are equal (one class), and the other nine have two classes. Table 1 lists representatives; the two-class orbits are numbered as in the scripts

(orbit_coord_ascent.py, class_flips.py), where the numbers 4 and 11–14 are the one-class orbits.

orbit	$2h$ on $+++,++-, \dots, ---$	size	classes	$\hat{h}(\{1,2,3\})$	status
A12	$(-2, -2, -2, 1, 0, -1, -1, -1)$	12	one	$\frac{1}{4}$	trivial (g_0 , Thm. E; Prop. 11.6)
A24	$(-2, -2, -1, 1, 0, 0, -1, -1)$	24	one	$\frac{1}{8}$	trivial (g_1 , Props. 9.10, 11.10)
D12	$(-1, -2, -2, 1, 0, -1, -1, 0)$	12	one	$\frac{1}{8}$	trivial (g_2 , Prop. 9.11)
B12	$(-2, -1, -1, 1, 0, 0, 0, -1)$	12	one	$\frac{1}{8}$	trivial (Prop. 11.15)
C12	$(-1, -2, -2, 1, -1, 0, 0, -1)$	12	one	$\frac{1}{8}$	trivial (Prop. 11.12)
rep1	$(-2, -2, -1, 0, -1, 1, 0, -1)$	24	$\{1\}, \{2, 3\}$	$\frac{1}{4}$	trivial (Prop. 11.8); equiv. to rep9
rep2	$(-2, -2, -1, 0, 1, -2, -1, -1)$	24	$\{1, 2\}, \{3\}$	$-\frac{1}{8}$	trivial (Prop. 11.6); implies A12
rep3	$(-2, -2, -1, 1, 0, -1, -1, 0)$	24	$\{1, 3\}, \{2\}$	0	trivial (Prop. 11.5)
rep5	$(-2, -1, -1, 0, 1, -2, 0, -1)$	24	$\{1, 2\}, \{3\}$	$-\frac{1}{8}$	open; implies D12
rep6	$(-2, -1, 0, -1, 1, -2, -1, 0)$	24	$\{1, 2\}, \{3\}$	$-\frac{1}{8}$	trivial (Prop. 11.12); implies C12
rep7	$(-1, -2, -1, 0, 1, -2, -1, 0)$	24	$\{1, 2\}, \{3\}$	$-\frac{1}{8}$	trivial (Prop. 11.10); implies A24
rep8	$(-1, -2, -1, 0, 1, -2, 0, -1)$	24	$\{1, 2\}, \{3\}$	0	trivial (Prop. 11.5)
rep9	$(-1, -2, 0, -1, 1, -2, -1, 0)$	24	$\{1, 2\}, \{3\}$	$-\frac{1}{4}$	trivial (Prop. 11.8); equiv. to rep1
rep10	$(-1, -2, 0, 0, 1, -1, -1, 0)$	24	$\{1, 2\}, \{3\}$	$-\frac{1}{8}$	open; implies B12

Table 1: The 14 orbits of boundary functions for $m = 5$. “Trivial” means $s_g = \text{triv}_g < s_g^U$ for every g with h in the orbit; “implies” and “equivalent” refer to Lemma 11.3 (“rep10 implies B12”: if rep10 is trivial then so is B12).

A further operation acts on the class-wise criterion (c) of Theorem D. For a class K of h let $h^K(w) = h(w^K)$, where w^K is w with the coordinates in K negated.

Lemma 11.3 (class flips). (a) If $-\hat{h}(\{j\})$ ($j \in K$) differs from every $\hat{h}(\{j'\})$ with $j' \notin K$, then h^K has the same classes as h , and condition (c) of Theorem D holds for h^K iff it holds for h . (b) In general, if condition (c) holds for h^K then it holds for h ; that is, h trivial implies h^K trivial.

Proof. The Fourier coefficients of h^K are $(-1)^{|S \cap K|} \hat{h}(S)$, so μ and σ are unchanged and the level-one coefficients in K change sign. Moreover $\pi_b^K(w) = \pi_b(w^K)$, so a label matrix C^K for h^K has the law of C with the columns in K negated, and the multilinear extension of h^K is $y \mapsto h(y^K)$. Let F be class-wise for h^K and write $F^{(j)}$ for the decoder of the h^K -class of j . Define decoders for h by $G_c = F^{(j)}$ for a class $c \neq K$ of h and any $j \in c$ (well defined, since the classes of h other than K are contained in classes of h^K), and $G_K(z) = -F^{(j)}(-z)$ for $j \in K$. Then $h^K(F^{(j)}(C_j^K))_j$ has the law of $h((-F^{(j)}(-C_j))_{j \in K}, (F^{(j)}(C_j))_{j \notin K}) = h(G_{c(j)}(C_j))_j$, so every value of Γ_ν for h^K is a value of Γ_ν for h . This proves (b). Under the hypothesis of (a) the classes coincide and $F \mapsto G$ is a bijection, which gives (a). $\square$

By `class_flips.py`, flipping either class of rep1 (rep3) gives a function in the orbit of rep9 (rep8) with the classes kept, and flipping either class of rep2, rep5, rep6, rep7 and rep10 merges the two classes and gives a function in the orbit of A12, D12, C12, A24 and B12, respectively. This gives the last column of Table 1. So, up to the equivalences $\text{rep1} \iff \text{rep9}$ and $\text{rep3} \iff \text{rep8}$, the 240 functions outside the orbits of $h_{g_0}, h_{g_1}, h_{g_2}$ form nine problems: B12, C12, and the two-class orbits rep1, rep2, rep3, rep5, rep6, rep7, rep10, where a solution of rep10 (rep6) would also settle B12 (C12). Propositions 11.5, 11.6, 11.8, 11.10 and 11.12 below solve rep3, rep2, rep1, rep7 and rep6 (and with it C12), and Proposition 11.15 solves B12, which leaves two: rep5 and rep10. In the merging case the one-class problem for h^K is the two-class problem for h restricted to decoders with $G_K(z) = -G_c(-z)$, where c is the other class; so each of rep2, rep5, rep6, rep7 and rep10 is at least as hard as the one-class orbit it implies.

Proposition 11.4 (computer-assisted). *Let h be one of the 240 boundary functions outside the orbits of $h_{g_0}, h_{g_1}, h_{g_2}$ (the orbits B12, C12 and the nine two-class orbits of Table 1). Then $\sigma\Gamma_\nu(F) \geq 0$ for every $k \leq 3$, every $\nu \in \{-1, 1\}^k$ and every class-wise decoder F with values in $[-1, 1]$. So none of these h has a certificate as in Theorem D(c) (or (d)) with at most three label bits.*

Proof. The three operations of Lemma 11.2 map classes to classes, and its proof applies verbatim to class-wise decoders (apply $F \mapsto \tilde{F}$ to the decoder of each class); none of them changes k . So it suffices to treat one function in each orbit. For B12 and C12 we use Lemma 11.3: by its proof, for a class K of rep10 (rep6) every value of Γ_ν for h^K , which lies in the orbit of B12 (C12), is a value of Γ_ν for rep10 (rep6) with the same k and ν . It therefore suffices to treat the nine two-class representatives of Table 1. All of them have $\mu < 0$, so we must show $\Gamma_\nu(F) \leq 0$.

Let $\{i, j\}$ be the pair class and s the singleton class, with decoders F_p and F_s . Since h is multilinear and only column s uses F_s , Γ_ν is affine in each value $F_s(z)$; for fixed F_p its maximum over F_s is therefore attained at a vertex of $[-1, 1]^{\{-1, 1\}^k}$. For F_s fixed at a vertex, Γ_ν is a quadratic polynomial in the 2^k values of F_p , as only the columns i and j use F_p . The maximum of a quadratic polynomial over a cube is attained at a vertex or at a critical point of its restriction to the relative interior of a face on which its Hessian block is nonsingular: if the maximum over a face is attained at a relative interior point and the block is singular, the polynomial is affine, hence constant, along a kernel direction through that point, so the maximum is also attained on the relative boundary of the face. All these critical points are rational. A permutation τ of the label rows, applied also to the label bits of both decoders, maps $\Gamma_\nu(F)$ to $\Gamma_{\nu \circ \tau}$ of the permuted decoder, so ν and $\nu \circ \tau$ give the same set of values and we may take $\nu = (1, \dots, 1, -1, \dots, -1)$. For each $k \leq 3$, each such ν and each vertex F_s , the script `orbit_k2_exact.py` ($k \leq 2$, Python fractions) and `orbit_k3_exact.py` ($k = 3$, python-flint; for each set of free coordinates the Hessian block is inverted once) enumerate all faces of the cube of values of F_p and solve the linear systems in exact rational arithmetic. In every case the maximum of Γ_ν is at most 0: it is $-\frac{1}{3}$ for rep2 with $\nu = (-1, \dots, -1)$ and 0 in all other cases (outputs `orbit_k2_exact.txt`, `orbit_k3_exact.txt`; the two scripts agree vertex by vertex for $k \leq 2$). $\square$

The value 0 is attained for every $k \geq 1$, so Proposition 11.4 is sharp: for the dictator decoder ($k = 1$, $\nu_1 = -\sigma$, $F(z) = z$ for every class) $\Gamma_\nu = \mathbb{E}_{\pi_{-\sigma}} h = \frac{\mu - \sigma \|h\|_2^2}{1 - \sigma\mu} = 0$, since $\|h\|_2^2 = |\mu|$ (decoders that ignore further label rows extend this to every $k \geq 1$). It upgrades the check over decoders with values in $\{-1, 0, 1\}$ above to all $[-1, 1]$ -valued decoders, and from $k \leq 2$ to $k \leq 3$; for $k \geq 4$ the exhaustive enumeration is out of reach (2^{16} vertices F_s and 3^{16} faces each).

By Theorem E, Propositions 9.10 and 9.11 and Lemma 11.2, every g as in Theorem B whose h lies in the orbit of h_{g_0} (12 functions), of h_{g_1} (24 functions) or of h_{g_2} (12 functions) has $s_g = \text{triv}_g < s_g^U$. For g_0 itself the dictator decoder ($k = 1$, $\nu = +1$, $F(z) = z$) gives exactly $\Gamma_+ = 0$, and it is a strict local maximum along the admissible directions ($F(z) = z(1 - \epsilon_z)$, $\epsilon_\pm \geq 0$; `q6_g0_variation.py`); coordinate ascent (exact one-dimensional steps, floating point) for every $k \leq 6$ finds the maximum 0 whenever ν has a $+$ entry (`g0_coord_ascent.py`). Theorem E and Propositions 9.10 and 9.11 turn such numerical evidence into proofs for three orbits. Two two-class orbits yield to a different argument: their h has no cubic term, and raising it to 0 off the support of π_+ gives an upper bound h' that also lacks one of the three pair terms; so $\mathbb{E}[h'(Y)]$ depends on the rows only through two of their three pair marginals, and the rows may be re-coupled without changing these.

Proposition 11.5. *Let g be as in Theorem B with h in the orbit of rep3 or of rep8 (Table 1; 48 functions). Then $s_g^L = s_g = \text{triv}_g < s_g^U$.*

Proof. For these h , $|\mu| = \frac{3}{8}$ and σh takes the value $-\frac{1}{2}$, where $\sigma = \text{sgn } \mu$, so $s_g^U > \text{triv}_g$ by Theorem C(ii). By Lemma 11.2, whose proof applies verbatim to class-wise decoders, and by Lemma 11.3(a) (the level-one coefficients of rep3 are $-\frac{1}{8}, -\frac{1}{4}, -\frac{1}{8}$, so flipping the class $\{2\}$ keeps the classes, and the flipped function is rep8 after a permutation of the coordinates: $h_{\text{rep8}}(w_1, w_2, w_3) = h_{\text{rep3}}(w_2, -w_3, w_1)$), it suffices to show that condition (c) of Theorem D fails for the representative h of rep3 , that is, $\Gamma_\nu(F) \leq 0$ for all k, ν and class-wise F . Then $s_g = \text{triv}_g$ by Theorem D, and $\text{triv}_g \leq s_g^L \leq s_g$.

In the order $+++ , ++- , \dots , ---$, $\pi_+ = \frac{1}{10}(0, 0, 1, 3, 2, 1, 1, 2)$ and $\pi_- = \frac{1}{22}(4, 4, 3, 1, 2, 3, 3, 2)$. Write X, V, Z for the columns 1, 2, 3 of C , F for the decoder of the class $\{1, 3\}$ and F' for that of $\{2\}$, and $Y = (F(X), F'(V), F(Z))$, so that $\Gamma_\nu(F) = \mathbb{E}[h(Y)]$. Let h' be h raised to 0 at $+++$ and $++-$, the two vertices outside the support of π_+ . Then $h' - h$ is multilinear and nonnegative at the vertices, hence on $[-1, 1]^3$, so $\Gamma_\nu(F) \leq \mathbb{E}[h'(Y)]$, and

$$2h'(y) = -\frac{1}{4} + \frac{1}{4}(y_1 - y_3) - \frac{1}{4}y_1y_3 + \frac{1}{2}y_2y_3.$$

The columns X and Z have the same law, since $\pi_b(w_1 = 1) = \pi_b(w_3 = 1)$ for both b , so $\mathbb{E}F(X) = \mathbb{E}F(Z)$ and

$$8\mathbb{E}[h'(Y)] = -1 - \mathbb{E}[F(X)F(Z)] + 2\mathbb{E}[F'(V)F(Z)]. \quad (20)$$

The right side depends on the law of C only through the laws of (X, Z) and (V, Z) , that is, through the marginals of the row laws π_{ν_i} on the coordinates $\{1, 3\}$ and $\{2, 3\}$. Put

$$\tilde{\pi}_b(w) = \pi_b(w) + c_b w_1 w_2, \quad c_+ = \frac{1}{10}, \quad c_- = -\frac{1}{22};$$

explicitly $\tilde{\pi}_+ = \frac{1}{10}(1, 1, 0, 2, 1, 0, 2, 3)$ and $\tilde{\pi}_- = \frac{1}{22}(3, 3, 4, 2, 3, 4, 2, 1)$, which are probability distributions. Summing $w_1 w_2$ over w_2 for fixed (w_1, w_3) , or over w_1 for fixed (w_2, w_3) , gives 0, so $\tilde{\pi}_b$ has the same marginals as π_b on $\{1, 3\}$ and on $\{2, 3\}$. Moreover $\mathbb{E}_{\tilde{\pi}_b}[w_1 w_2] = \mathbb{E}_{\pi_b}[w_1 w_2] + 8c_b = \mathbb{E}_{\pi_b}[w_2 w_3]$ (the values are $-\frac{2}{5} + \frac{4}{5} = \frac{2}{5}$ for $b = +1$ and $\frac{2}{11} - \frac{4}{11} = -\frac{2}{11}$ for $b = -1$), and w_1 and w_3 have the same law, so under $\tilde{\pi}_b$ the pair (w_2, w_1) has the same law as (w_2, w_3) . Let $\tilde{C}$ have independent rows, row $i \sim \tilde{\pi}_{\nu_i}$, and columns $\tilde{X}, \tilde{V}, \tilde{Z}$. As the rows are independent, $(\tilde{X}, \tilde{Z})$ and $(\tilde{V}, \tilde{Z})$ have the laws of (X, Z) and (V, Z) (the law of a pair of ± 1 -valued variables is determined by their means and the mean of their product), so the right side of (20) is unchanged when (X, V, Z) is replaced by $(\tilde{X}, \tilde{V}, \tilde{Z})$; moreover $(\tilde{V}, \tilde{X})$ has the same law as $(\tilde{V}, \tilde{Z})$, so $\mathbb{E}[F'(\tilde{V})F(\tilde{Z})] = \frac{1}{2}\mathbb{E}[F'(\tilde{V})(F(\tilde{X}) + F(\tilde{Z}))]$. Hence

$$\begin{aligned} 8\mathbb{E}[h'(Y)] &= -\mathbb{E}[\phi(F(\tilde{X}), F(\tilde{Z}), F'(\tilde{V}))], \\ \phi(a, b, t) &= 1 + ab - t(a + b) = \frac{1+t}{2}(1-a)(1-b) + \frac{1-t}{2}(1+a)(1+b), \end{aligned}$$

and $\phi \geq 0$ on $[-1, 1]^3$. So $\Gamma_\nu(F) \leq \mathbb{E}[h'(Y)] \leq 0$. $\square$

In terms of $P = \frac{1+F}{2}$ and $P' = \frac{1+F'}{2}$ the proof gives $\Gamma_\nu(F) \leq -\frac{1}{2}\mathbb{E}[P'(\tilde{V})(1-P(\tilde{X}))(1-P(\tilde{Z})) + (1-P'(\tilde{V}))P(\tilde{X})P(\tilde{Z})]$. This certificate is not of the level-weight type of Theorem E: it writes the (quadratic) integrand as a sum of two nonnegative cubic terms in the decoder values, evaluated under the joint law of three columns of $\tilde{C}$, whose within-row coupling $\tilde{\pi}_b$ differs from π_b . For the dictator decoder on a row with $\nu_r = +1$ (with $P'(z) = \mathbf{1}\{z_r = 1\}$) both terms vanish, because $\tilde{\pi}_+(+-+) = \tilde{\pi}_+(-+-) = 0$. With three columns the same device settles no other two-class orbit: for each of the other seven two-class orbits, no cubic-free $h' \geq h$ that agrees with h on the support of π_+ becomes pointwise nonpositive after symmetrising the pair class by a re-coupling that keeps the pair marginals entering h' . For each choice of the set of pair coefficients of h' that vanish, the conditions on h' form a linear system in at most three unknowns, and the re-coupled row laws

form an explicit interval family; both are decided exactly (`coupling_search.py`). For rep5 and rep7 suitable h' exist, but the required re-coupled row law of sign +1 does not (for the one-class orbits B12, C12 no admissible h' works, see the paragraph after Conjecture 11.14; for the one-class orbit A24 the device works, see the paragraph “One-class designs” after Proposition 11.12). The script `rep3_coupling.py` checks every number in the proof in exact arithmetic, and tests the chain $\Gamma_\nu(F) \leq \mathbb{E}[h'(Y)] = -\frac{1}{8}\mathbb{E}[\phi] \leq 0$ exactly on 1360 random decoders with $k \leq 3$.

With more columns the device does settle a further orbit. Instead of re-coupling the three columns of C , one builds a matrix with four columns, three of which carry the decoder of the pair class and one that of the singleton class, and whose pairs of columns realise the pair laws that enter h' .

Proposition 11.6. *Let g be as in Theorem B with h in the orbit of rep2 (Table 1; 24 functions). Then $s_g^L = s_g = \text{triv}_g < s_g^U$.*

Proof. For these h , $|\mu| = \frac{1}{2}$ and σh takes the value $-\frac{1}{2}$, where $\sigma = \text{sgn } \mu$, so $s_g^U > \text{triv}_g$ by Theorem C(ii). By Lemma 11.2 (whose proof applies verbatim to class-wise decoders) it suffices to show that $\Gamma_\nu(F) \leq 0$ for the representative h of rep2, for all k, ν and class-wise F ; then $s_g = \text{triv}_g$ by Theorem D, and $\text{triv}_g \leq s_g^L \leq s_g$.

In the order $+++ , ++- , \dots , ---$ of the vertices,

$$2h = (-2, -2, -1, 0, 1, -2, -1, -1), \quad \pi_+ = \frac{1}{8}(0, 0, 1, 2, 3, 0, 1, 1), \quad \pi_- = \frac{1}{24}(4, 4, 3, 2, 1, 4, 3, 3).$$

The level-one coefficients are $-\frac{1}{8}, -\frac{1}{8}, \frac{1}{8}$, so the classes are $\{1, 2\}$ and $\{3\}$. Write X_1, X_2, X_3 for the columns of C , F for the decoder of $\{1, 2\}$, F' for that of $\{3\}$, and $Y = (F(X_1), F(X_2), F'(X_3))$. Let h' be h raised by $\frac{1}{2}$ at $+++$ and $++-$ and by 1 at $-+-$, the three vertices outside the support of π_+ (so h' takes the values $-\frac{1}{2}, -\frac{1}{2}, 0$ there). Then $h' - h$ is multilinear and nonnegative at the vertices, hence on $[-1, 1]^3$, so $\Gamma_\nu(F) \leq \mathbb{E}[h'(Y)]$, and

$$8h'(y) = -2 - (y_1 - y_2) - 2y_1y_2 - (y_1 - y_2)y_3.$$

Since $\pi_b(w_1 = 1) = \pi_b(w_2 = 1)$ for both b , the columns X_1 and X_2 have the same law, so $\mathbb{E}F(X_1) = \mathbb{E}F(X_2)$ and

$$8\mathbb{E}[h'(Y)] = -2 - 2\mathbb{E}[F(X_1)F(X_2)] - \mathbb{E}[F(X_1)F'(X_3)] + \mathbb{E}[F(X_2)F'(X_3)]. \quad (21)$$

The right side depends on the law of C only through the laws of the three pairs of columns. Under π_+ the pairs (w_1, w_2) , (w_1, w_3) , (w_2, w_3) take the values $++ , +- , -+ , --$ with probabilities $(0, \frac{3}{8}, \frac{3}{8}, \frac{1}{4})$, $(\frac{1}{8}, \frac{1}{4}, \frac{1}{2}, \frac{1}{8})$ and $(\frac{3}{8}, 0, \frac{1}{4}, \frac{3}{8})$. Let ρ_+ be the law on $\{-1, 1\}^4$ with

$$\rho_+(- - + +) = \rho_+(- + - +) = \rho_+(+ - - -) = \frac{1}{4}, \quad \rho_+(- - + -) = \rho_+(+ + - +) = \frac{1}{8},$$

and let $\rho_- = \frac{1}{12}(1 - 4\rho_+)$, which is 0 at the three points of ρ_+ -mass $\frac{1}{4}$, $\frac{1}{24}$ at the two points of mass $\frac{1}{8}$ and $\frac{1}{12}$ at the other eleven points of $\{-1, 1\}^4$; both are probability distributions. Counting atoms, under ρ_+ the pairs (u_1, u_3) and (u_2, u_3) have the law of (w_1, w_2) under π_+ , the pair (u_1, u_4) that of (w_1, w_3) and the pair (u_2, u_4) that of (w_2, w_3) . The same holds for ρ_- and π_- : since $(1 + \mu)\pi_+ + (1 - \mu)\pi_- = \frac{1}{4}$ on $\{-1, 1\}^3$ and $(1 + \mu)\rho_+ + (1 - \mu)\rho_- = \frac{1}{8}$ on $\{-1, 1\}^4$, every pair marginal $L_\pm$ of $\pi_\pm$, and of $\rho_\pm$, satisfies $(1 + \mu)L_+(x, y) + (1 - \mu)L_-(x, y) = \frac{1}{2}$, so L_- is determined by L_+ . Let U have independent rows, row $i \sim \rho_{\nu_i}$, and columns A, B, D, E . Then (A, D) and (B, D) have the law of (X_1, X_2) , (A, E) has the law of (X_1, X_3) and (B, E) that of (X_2, X_3) . By (21),

$$\begin{aligned} 8\mathbb{E}[h'(Y)] &= -2 - \mathbb{E}[F(D)(F(A) + F(B))] - \mathbb{E}[F'(E)(F(A) - F(B))] \\ &= -\mathbb{E}[\psi(F(A), F(B), F(D), F'(E))], \end{aligned}$$

where

$$\begin{aligned}\psi(a, b, d, e) &= 2 + d(a + b) + e(a - b) \\ &= \frac{1}{2}[(1 + a)(1 + b)(1 + d) + (1 - a)(1 - b)(1 - d)] \\ &\quad + \frac{1}{2}[(1 + a)(1 - b)(1 + e) + (1 - a)(1 + b)(1 - e)] \geq 0\end{aligned}$$

on $[-1, 1]^4$. So $\Gamma_\nu(F) \leq \mathbb{E}[h'(Y)] \leq 0$. □

Remark 11.7. (a) In terms of $P = \frac{1+F}{2}$ and $P' = \frac{1+F'}{2}$ the proof gives

$$\begin{aligned}\Gamma_\nu(F) \leq & -\frac{1}{2}\mathbb{E}[P(A)P(B)P(D) + (1 - P(A))(1 - P(B))(1 - P(D)) \\ & + P(A)(1 - P(B))P'(E) + (1 - P(A))P(B)(1 - P'(E))].\end{aligned}$$

For the dictator decoder on a row with $\nu_r = +1$ all four terms vanish, because no atom of ρ_+ has $u_1 = u_2 = u_3$ or $(u_1, u_2, u_4) = \pm(+, -, +)$. Pair laws alone cannot do this with three columns: the three-column version of the device fails for rep2 (`coupling_search.py`).

(b) Flipping the class $\{3\}$ of rep2 merges the classes and gives a function in the orbit of h_{g_0} : for the representative, $h(w_1, w_2, -w_3) = h_{g_0}(w_2, w_1, w_3)$ at all eight vertices. So by Lemma 11.3(b) and Lemma 11.2, Proposition 11.6 implies Proposition 9.1 and hence Theorem E. This second proof of Theorem E uses neither Theorem K nor any computation; Proposition 11.10 below does the same for g_1 .

(c) The script `rep2_coupling.py` checks every number in the proof in exact arithmetic (including the flip to the orbit of h_{g_0}), and tests the chain $\Gamma_\nu(F) \leq \mathbb{E}[h'(Y)] = -\frac{1}{8}\mathbb{E}[\psi] \leq 0$ exactly on 1360 random decoders with $k \leq 3$.

If h' keeps a cubic term, the law of C enters through the full row laws. The device still applies if three of the new columns carry an exact copy of the rows, and the other pairs of columns realise pair laws as before. This settles two more orbits.

Proposition 11.8. *Let g be as in Theorem B with h in the orbit of rep1 or of rep9 (Table 1; 48 functions). Then $s_g^L = s_g = \text{triv}_g < s_g^U$.*

Proof. For these h , $|\mu| = \frac{3}{8}$ and σh takes the value $-\frac{1}{2}$, where $\sigma = \text{sgn } \mu$, so $s_g^U > \text{triv}_g$ by Theorem C(ii). By Lemma 11.2 (whose proof applies verbatim to class-wise decoders) and by Lemma 11.3(a) (the level-one coefficients of rep1 are $-\frac{1}{4}, -\frac{1}{8}, -\frac{1}{8}$, so flipping the class $\{1\}$ keeps the classes, and the flipped function is rep9 after a permutation of the coordinates: $h_{\text{rep9}}(w_1, w_2, w_3) = h_{\text{rep1}}(-w_3, w_2, w_1)$), it suffices to show that $\Gamma_\nu(F) \leq 0$ for the representative h of rep1, for all k, ν and class-wise F ; then $s_g = \text{triv}_g$ by Theorem D, and $\text{triv}_g \leq s_g^L \leq s_g$.

In the order $+++ , ++- , \dots , ---$ of the vertices,

$$2h = (-2, -2, -1, 0, -1, 1, 0, -1), \quad \pi_+ = \frac{1}{10}(0, 0, 1, 2, 1, 3, 2, 1), \quad \pi_- = \frac{1}{22}(4, 4, 3, 2, 3, 1, 2, 3).$$

The level-one coefficients are $-\frac{1}{4}, -\frac{1}{8}, -\frac{1}{8}$, so the classes are $\{1\}$ and $\{2, 3\}$. Write V, X, Z for the columns 1, 2, 3 of C , F' for the decoder of $\{1\}$, F for that of $\{2, 3\}$, and $Y = (F'(V), F(X), F(Z))$. Let h' be h raised by 1 at $++-$, one of the two vertices outside the support of π_+ , and equal to h elsewhere. Then $\Gamma_\nu(F) \leq \mathbb{E}[h'(Y)]$ as before, and

$$8h'(y) = -2 - y_1 - 2y_3 - y_1y_2 - y_1y_3 - 2y_2y_3 + y_1y_2y_3.$$

Since $\pi_b(w_2 = 1) = \pi_b(w_3 = 1)$ for both b , $\mathbb{E}F(X) = \mathbb{E}F(Z)$ and

$$\begin{aligned} 8\mathbb{E}[h'(Y)] = & -2 - \mathbb{E}F'(V) - 2\mathbb{E}F(X) - \mathbb{E}[F'(V)F(X)] \\ & - \mathbb{E}[F'(V)F(Z)] - 2\mathbb{E}[F(X)F(Z)] + \mathbb{E}[F'(V)F(X)F(Z)]. \end{aligned} \quad (22)$$

Under π_+ the pairs (w_2, w_3) and (w_3, w_1) take the values $++, +-, -+, --$ with probabilities $(\frac{1}{10}, \frac{3}{10}, \frac{3}{10}, \frac{3}{10})$ and $(\frac{1}{10}, \frac{3}{10}, \frac{1}{5}, \frac{2}{5})$. Let ρ_+ be the law on $\{-1, 1\}^4$, with coordinates (x, z, z', v) , given by

$$\begin{aligned} \rho_+(+ - - -) &= \frac{1}{5}, & \rho_+(+ + - -) &= \rho_+(+ - + -) = \rho_+(- + + -) = \rho_+(- + - +) = \frac{1}{10}, \\ \rho_+(- + - -) &= \rho_+(- - + +) = \rho_+(- - + -) = \rho_+(- - - +) = \frac{1}{10}, \end{aligned}$$

and let $\rho_- = \frac{1}{11}(1 - 5\rho_+)$, which is 0 at $+ - - -$, $\frac{1}{22}$ at the eight points of ρ_+ -mass $\frac{1}{10}$ and $\frac{1}{11}$ at the other seven points of $\{-1, 1\}^4$. Counting atoms, under ρ_+ the triple (x, z, v) has the law of (w_2, w_3, w_1) under π_+ , the pairs (x, z') and (z, z') have the law of (w_2, w_3) , and (z', v) has that of (w_3, w_1) . The same holds for ρ_- and π_- : since $(1 + \mu)\pi_+ + (1 - \mu)\pi_- = \frac{1}{4}$ on $\{-1, 1\}^3$ and $(1 + \mu)\rho_+ + (1 - \mu)\rho_- = \frac{1}{8}$ on $\{-1, 1\}^4$, every marginal $L_\pm$ of $\pi_\pm$ or of $\rho_\pm$ on $t \in \{2, 3\}$ coordinates satisfies $(1 + \mu)L_+ + (1 - \mu)L_- = 2^{1-t}$, so L_- is determined by L_+ . Let U have independent rows, row $i \sim \rho_{\nu_i}$, and columns $\hat{X}, \hat{Z}, \hat{Z}', \hat{V}$. Then $(\hat{X}, \hat{Z}, \hat{V})$ has the law of (X, Z, V) , $(\hat{X}, \hat{Z}')$ and $(\hat{Z}, \hat{Z}')$ have the law of (X, Z) , and $(\hat{Z}', \hat{V})$ has the law of (Z, V) . So $\mathbb{E}F'(V) = \mathbb{E}F'(\hat{V})$, $\mathbb{E}F(X) = \mathbb{E}F(\hat{X}) = \mathbb{E}F(\hat{Z}')$, $2\mathbb{E}[F(X)F(Z)] = \mathbb{E}[F(\hat{X})F(\hat{Z}')] + \mathbb{E}[F(\hat{Z})F(\hat{Z}')]$, $\mathbb{E}[F'(V)F(X)] = \mathbb{E}[F'(\hat{V})F(\hat{X})]$, $\mathbb{E}[F'(V)F(Z)] = \mathbb{E}[F'(\hat{V})F(\hat{Z}')] + \mathbb{E}[F'(\hat{V})F(\hat{Z})]$, and (22) becomes $\mathbb{E}[h'(Y)] = \mathbb{E}[R(F(\hat{X}), F(\hat{Z}), F(\hat{Z}'), F'(\hat{V}))]$ with

$$8R(x, z, z', v) = -2 - x - z' - v - xv - xz' - zz' - z'v + xzv.$$

Finally

$$\begin{aligned} -8R &= (1+x)(1+z')(1+v) + \frac{1}{4}[(1+x)(1+z)(1+z')(1-v) + (1+x)(1-z)(1-z')(1+v) \\ &\quad + (1-x)(1+z)(1+z')(1+v) + (1-x)(1-z)(1-z')(1-v)], \end{aligned}$$

which is nonnegative on $[-1, 1]^4$. So $\Gamma_\nu(F) \leq \mathbb{E}[h'(Y)] \leq 0$. $\square$

Remark 11.9. (a) In terms of $P = \frac{1+F}{2}$ and $P' = \frac{1+F'}{2}$, and writing $P_X = P(\hat{X})$ and so on, the proof gives

$$\begin{aligned} \Gamma_\nu(F) \leq & -\mathbb{E}[P_X P_{Z'} P'_V] - \frac{1}{2}\mathbb{E}[P_X P_Z P_{Z'}(1 - P'_V) + P_X(1 - P_Z)(1 - P_{Z'})P'_V \\ & + (1 - P_X)P_Z P_{Z'} P'_V + (1 - P_X)(1 - P_Z)(1 - P_{Z'})(1 - P'_V)]. \end{aligned}$$

The polynomial R vanishes at every atom of ρ_+ , as it must, since the dictator decoder on a row with $\nu_r = +1$ has $\Gamma_\nu = 0$. (b) The script `rep1_coupling.py` checks every number in the proof in exact arithmetic (including the two class flips to the orbit of rep9), and tests the chain $\Gamma_\nu(F) \leq \mathbb{E}[h'(Y)] = \mathbb{E}[R] \leq 0$ exactly on 1360 random decoders with $k \leq 3$.

With two columns for the singleton class as well, the same idea settles rep7, and here no majorant h' is needed.

Proposition 11.10. *Let g be as in Theorem B with h in the orbit of rep7 (Table 1; 24 functions). Then $s_g^L = s_g = \text{triv}_g < s_g^U$.*

Proof. For these h , $|\mu| = \frac{3}{8}$ and σh takes the value $-\frac{1}{2}$, where $\sigma = \text{sgn } \mu$, so $s_g^U > \text{triv}_g$ by Theorem C(ii). By Lemma 11.2 (whose proof applies verbatim to class-wise decoders) it suffices to show that $\Gamma_\nu(F) \leq 0$ for the representative h of rep7, for all k, ν and class-wise F ; then $s_g = \text{triv}_g$ by Theorem D, and $\text{triv}_g \leq s_g^L \leq s_g$.

In the order $+++ , ++- , \dots , ---$ of the vertices,

$$2h = (-1, -2, -1, 0, 1, -2, -1, 0), \quad \pi_+ = \frac{1}{10}(1, 0, 1, 2, 3, 0, 1, 2), \quad \pi_- = \frac{1}{22}(3, 4, 3, 2, 1, 4, 3, 2),$$

$$8h(y) = -3 - y_1 - y_2 + y_3 - y_1y_2 - y_1y_3 + 3y_2y_3 - y_1y_2y_3,$$

so the classes are $\{1, 2\}$ and $\{3\}$. Write X_1, X_2, X_3 for the columns of C , F for the decoder of $\{1, 2\}$ and F' for that of $\{3\}$. Since X_1 and X_2 have the same law,

$$8\Gamma_\nu(F) = -3 - 2\mathbb{E}F(X_1) + \mathbb{E}F'(X_3) - \mathbb{E}[F(X_1)F(X_2)] - \mathbb{E}[F(X_1)F'(X_3)] \\ + 3\mathbb{E}[F(X_2)F'(X_3)] - \mathbb{E}[F(X_1)F(X_2)F'(X_3)]. \quad (23)$$

Under π_+ the pairs $(w_1, w_2), (w_1, w_3), (w_2, w_3)$ take the values $++, +-, -+, --$ with probabilities $(\frac{1}{10}, \frac{3}{10}, \frac{3}{10}, \frac{3}{10}), (\frac{1}{5}, \frac{1}{5}, \frac{2}{5}, \frac{1}{5})$ and $(\frac{2}{5}, 0, \frac{1}{5}, \frac{2}{5})$, and under π_- with probabilities $\frac{1}{22}(7, 5, 5, 5), \frac{1}{22}(6, 6, 4, 6)$ and $\frac{1}{22}(4, 8, 6, 4)$. Let ρ_+ and ρ_- be the laws on $\{-1, 1\}^5$, with coordinates (a, b, c, d, e) , given by

$$\rho_+ = \frac{1}{5}(\delta_{++++} + \delta_{--++} + \delta_{----}) + \frac{1}{10}(\delta_{+-++} + \delta_{+---} + \delta_{-+++} + \delta_{-+-+}), \\ \rho_- = \frac{2}{11}(\delta_{+++-} + \delta_{+-+-}) + \frac{3}{22}(\delta_{-++-} + \delta_{----}) \\ + \frac{1}{11}(\delta_{-++++} + \delta_{--++-} + \delta_{----+}) + \frac{1}{22}(\delta_{+-++-} + \delta_{+---+}),$$

where δ_u is the unit mass at u . Counting atoms, under ρ_b the triple (a, c, e) has the law of (w_2, w_1, w_3) under π_b , the pair (b, c) has the law of (w_1, w_2) , the pair (a, d) that of (w_1, w_3) , and the pairs (b, d) and (b, e) that of (w_2, w_3) . Let U have independent rows, row $i \sim \rho_{\nu_i}$, and columns A_1, A_2, A_3, B_1, B_2 . Then (A_1, A_3, B_2) has the law of (X_2, X_1, X_3) , (A_2, A_3) that of (X_1, X_2) , (A_1, B_1) that of (X_1, X_3) , and (A_2, B_1) and (A_2, B_2) that of (X_2, X_3) ; also (A_1, B_2) , a pair inside the triple, has the law of (X_2, X_3) , and A_1, A_2 have the law of X_1 . Matching the terms of (23) accordingly,

$$\Gamma_\nu(F) = \mathbb{E}[R(F(A_1), F(A_2), F(A_3), F'(B_1), F'(B_2))], \\ 8R(a, b, c, d, e) = -3 - a - b + e - bc - ad + ae + bd + be - ace.$$

As R is multilinear, it suffices to show $R \leq 0$ at the vertices. For fixed (a, b, c) the minimum of $-8R = 3 + a + b + bc + d(a - b) + e(ac - 1 - a - b)$ over $d, e \in \{-1, 1\}$ is $3 + a + b + bc - |a - b| - |ac - 1 - a - b|$, which equals $2 + 2c$ if $a = b = 1$ and 0 at the six other vertices (a, b, c) . So $R \leq 0$ on $[-1, 1]^5$ and $\Gamma_\nu(F) \leq 0$. $\square$

Remark 11.11. (a) The identity $\Gamma_\nu(F) = \mathbb{E}[R]$ is exact: every coefficient of h is carried by the columns of U , and the cubic one by the copy (A_1, A_3, B_2) of the rows. The polynomial R vanishes at every atom of ρ_+ (dictator decoder). (b) Flipping the class $\{3\}$ of rep7 merges the classes and gives a function in the orbit of h_{g_1} : for the representative, $h(w_1, w_2, -w_3) = h_{g_1}(w_2, w_3, w_1)$. So by Lemmas 11.3(b) and 11.2, Proposition 11.10 implies the inequality $\Gamma_\nu \leq 0$ for g_1 in Proposition 9.10, which gives a second proof of that proposition without Theorem K. (c) The laws $\rho_\pm$ are vertices of the linear programmes defined by the marginal conditions (their supports, of sizes 7 and 9, index linearly independent columns of the constraint matrix). For $b = +1$ every admissible law has an atom of mass at least $\frac{1}{5}$, so $(1 + \mu)\rho_+ \leq 2^{1-N} = 2^{-4}$ fails, and the complementary construction $\rho_- = (2^{1-N} - (1 + \mu)\rho_+)/ (1 - \mu)$ used for rep2 and rep1 (with $N = 4$) is not available. The script `rep7_coupling.py` checks every number in the proof in exact arithmetic (including the flip to the orbit of h_{g_1}), and tests $\Gamma_\nu(F) = \mathbb{E}[R] \leq 0$ exactly on 1360 random decoders with $k \leq 3$.

With six columns that carry four copies of the rows, all sharing the singleton column, the device settles rep6, and with it the one-class orbit C12.

Proposition 11.12. *Let g be as in Theorem B with h in the orbit of rep6 or of C12 (Table 1; 36 functions). Then $s_g^L = s_g = \text{triv}_g < s_g^U$.*

Proof. For these h , $|\mu| = \frac{3}{8}$ and σh takes the value $-\frac{1}{2}$, where $\sigma = \text{sgn } \mu$, so $s_g^U > \text{triv}_g$ by Theorem C(ii). For the representative h of rep6 (below), flipping the class $\{3\}$ gives $h(w_1, w_2, -w_3) = h_C(w_2, w_1, w_3)$ at all eight vertices, where h_C is the representative of C12 in Table 1. By Lemmas 11.2 (whose proof applies verbatim to class-wise decoders) and 11.3(b) it therefore suffices to show that $\Gamma_\nu(F) \leq 0$ for the representative h of rep6, for all k, ν and class-wise F ; then $s_g = \text{triv}_g$ by Theorem D, and $\text{triv}_g \leq s_g^L \leq s_g$.

In the order $+++ , ++- , \dots , ---$ of the vertices,

$$2h = (-2, -1, 0, -1, 1, -2, -1, 0), \quad \pi_+ = \frac{1}{10}(0, 1, 2, 1, 3, 0, 1, 2), \quad \pi_- = \frac{1}{22}(4, 3, 2, 3, 1, 4, 3, 2),$$

$$8h(y) = -3 - y_1 - y_2 + y_3 - y_1y_2 - y_1y_3 + y_2y_3 - 3y_1y_2y_3,$$

so the classes are $\{1, 2\}$ and $\{3\}$. Write X_1, X_2, X_3 for the columns of C , F for the decoder of $\{1, 2\}$, F' for that of $\{3\}$, and $Y = (F(X_1), F(X_2), F'(X_3))$. Let h' be h raised by $\frac{1}{2}$ at $+++$ and $+-$, the two vertices outside the support of π_+ . Then $8(h' - h)(y) = (1 + y_2)(1 + y_1y_3) \geq 0$ on $[-1, 1]^3$, so $\Gamma_\nu(F) \leq \mathbb{E}[h'(Y)]$, and

$$8h'(y) = -2 - y_1 + y_3 - y_1y_2 + y_2y_3 - 2y_1y_2y_3,$$

so

$$8\mathbb{E}[h'(Y)] = -2 - \mathbb{E}F(X_1) + \mathbb{E}F'(X_3) - \mathbb{E}[F(X_1)F(X_2)] \\ + \mathbb{E}[F(X_2)F'(X_3)] - 2\mathbb{E}[F(X_1)F(X_2)F'(X_3)]. \quad (24)$$

Let ρ_+ and ρ_- be the laws on $\{-1, 1\}^6$, with coordinates $(a_1, \dots, a_5, e)$, given by

$$\rho_+ = \frac{1}{10}(\delta_{++++--} + \delta_{++----} + \delta_{+-+---} + \delta_{+---+-} + \delta_{-+++++} \\ + \delta_{-++---} + \delta_{-+----} + \delta_{--++--} + \delta_{--++--} + \delta_{-----}),$$

$$\rho_- = \frac{2}{11}\delta_{++++++} + \frac{3}{22}(\delta_{++++--} + \delta_{++--+-} + \delta_{-----}) \\ + \frac{1}{11}(\delta_{+-+---} + \delta_{-++---} + \delta_{-+---} + \delta_{-----}) + \frac{1}{22}\delta_{-+++++},$$

where δ_u is the unit mass at u ; so ρ_+ is uniform on ten points. Counting atoms, under ρ_b the triples (a_1, a_4, e) , (a_1, a_5, e) and (a_3, a_4, e) have the law of (w_1, w_2, w_3) under π_b , the triple (a_2, a_3, e) has the law of (w_2, w_1, w_3) , and the pairs (a_2, a_5) and (a_3, a_5) have the law of (w_1, w_2) . Let U have independent rows, row $i \sim \rho_{\nu_i}$, and columns $A_1, \dots, A_5, B$. Then (A_1, A_4, B) , (A_1, A_5, B) and (A_3, A_4, B) have the law of (X_1, X_2, X_3) , (A_2, A_3, B) has that of (X_2, X_1, X_3) , and (A_2, A_5) and (A_3, A_5) have that of (X_1, X_2) . Hence, as X_1 and X_2 have the same law, every A_t has the law of X_1 , the pairs (A_2, B) and (A_5, B) have the law of (X_2, X_3) , and $\mathbb{E}[F(A_t)F(A_u)] = \mathbb{E}[F(X_1)F(X_2)]$ for the six pairs $\{t, u\} = \{1, 4\}, \{1, 5\}, \{3, 4\}, \{2, 3\}, \{2, 5\}, \{3, 5\}$. Writing $2\mathbb{E}F(X_1) = \mathbb{E}F(A_2) + \mathbb{E}F(A_5)$, $2\mathbb{E}[F(X_2)F'(X_3)] = \mathbb{E}[F(A_2)F'(B)] + \mathbb{E}[F(A_5)F'(B)]$,

$$-2\mathbb{E}[F(X_1)F(X_2)] \\ = \mathbb{E}[F(A_1)F(A_4) + F(A_1)F(A_5) - F(A_2)F(A_3) + F(A_3)F(A_4) - 2F(A_2)F(A_5) - 2F(A_3)F(A_5)],$$

and $4 \mathbb{E}[F(X_1)F(X_2)F'(X_3)]$ as the sum of the four terms $\mathbb{E}[F(A_t)F(A_u)F'(B)]$ over the four triples, (24) becomes $\mathbb{E}[h'(Y)] = \mathbb{E}[R(F(A_1), \dots, F(A_5), F'(B))]$ with

$$16R(a, e) = -4 - a_2 - a_5 + 2e + a_1a_4 + a_1a_5 - a_2a_3 + a_3a_4 - 2a_2a_5 - 2a_3a_5 \\ + (a_2 + a_5)e - (a_1a_4 + a_1a_5 + a_2a_3 + a_3a_4)e.$$

As R is multilinear, it suffices to show $R \leq 0$ at the vertices. At $e = 1$, $16R = -2 - 2(a_2a_3 + a_2a_5 + a_3a_5) = 1 - (a_2 + a_3 + a_5)^2 \leq 0$, since $a_2 + a_3 + a_5$ is odd. At $e = -1$,

$$16R = -2[2 + (1 + a_2)(1 + a_5) + a_3(a_5 - a_4) - a_1(a_4 + a_5)] \leq 0,$$

since one of $a_5 - a_4$ and $a_4 + a_5$ vanishes and the other has modulus 2. So $R \leq 0$ on $[-1, 1]^6$ and $\Gamma_\nu(F) \leq \mathbb{E}[h'(Y)] \leq 0$. $\square$

Remark 11.13. (a) Here h' keeps a cubic term, which is carried by the four copies of the rows; they share the column B , and the same pair laws are realised by several pairs of columns. The polynomial R vanishes at every atom of ρ_+ (dictator decoder). Since $(1 + \mu)\rho_+ = \frac{1}{16} > 2^{-5}$ at the atoms of ρ_+ , the complementary construction of ρ_- used for rep2 and rep1 does not apply to this ρ_+ . A floating-point search finds no design of the type described below with at most five columns for rep6. (b) For C12 the method of Theorem E cannot work: every majorant without cubic term that agrees with h on the support of π_+ gives a functional that is positive for some decoder (see the discussion after Conjecture 11.14). Proposition 11.12 reaches C12 through the two-class problem rep6, which is at least as hard, with a majorant that keeps its cubic term. (c) The script `rep6_coupling.py` checks every number in the proof in exact arithmetic (including the coefficient matching and the flip to the orbit of C12), and tests the chain $\Gamma_\nu(F) \leq \mathbb{E}[h'(Y)] = \mathbb{E}[R] \leq 0$ exactly on 1360 random decoders with $k \leq 3$.

Coupling designs. The certificates of Propositions 11.5–11.12 are instances of one linear scheme. Choose N columns, each assigned to the pair class or to the singleton class; assign to some pairs of columns one of the three pair laws of π_b (compatible with the classes), and to some triples (two pair-class columns and one singleton column whose pairs carry the three different pair laws) the full row law. A certificate consists of laws $\rho_\pm$ on $\{-1, 1\}^N$ with the assigned marginals, a majorant $h' \geq h$ equal to h on the support of π_+ , and a multilinear $R \leq 0$ on $\{-1, 1\}^N$ whose monomials are supported on the assigned pairs and triples and whose coefficients, summed over the columns, pairs and triples carrying the same law, equal those of h' . Then $\Gamma_\nu(F) \leq \mathbb{E}[h'(Y)] = \mathbb{E}[R] \leq 0$ as in the proofs above. For a fixed assignment the existence of $\rho_\pm$ and of (h', R) are two linear feasibility problems. A floating-point search (`coupling_designs.py`) over all designs with at least two columns of the pair class and at least one of the singleton class, enumerated up to permutations of the columns within a class, gives the following. With $N \leq 5$ (1279 designs) it finds certificates for rep2, rep3 and rep8 with pair laws only ($N = 3$ for rep3 and rep8, $N = 4$ for rep2), with copies of the rows also for rep1 and rep9 ($N = 4$) and for rep7 ($N = 5$), and none for rep5, rep6 and rep10. With $N = 6$ (19,875 designs, copies of the rows allowed; run for rep5, rep6 and rep10 only) it finds 20 certificates for rep6, among them the design of Proposition 11.12, and none for rep5 and rep10. For rep10 no majorant without cubic term exists at all (its cubic coefficient is $-\frac{1}{8}$, and the only vertex outside the support of π_+ has $w_1w_2w_3 = -1$), so pair laws alone cannot work there. For a fixed number N of columns and a fixed assignment of the columns to the classes, the search over all assignments of laws to pairs and triples is a single mixed-integer linear feasibility problem: binary variables select the pair laws and the copies, the marginal conditions on $\rho_\pm$ are imposed for the selected pairs and triples, and the coefficient of R on a pair or triple

of columns is bounded in modulus by $|\mu|$ if the pair or triple is selected and vanishes otherwise. The bound $|\mu|$ is no restriction: $R \leq 0$ on $\{-1, 1\}^N$ gives $|\hat{R}(S)| \leq \mathbb{E}|R| = -\hat{R}(\emptyset) = -\hat{h}'(\emptyset) \leq -\mu$ for every S (expectations over the uniform measure on $\{-1, 1\}^N$), as $h' \geq h$. A floating-point solution of these problems (`milp_designs.py`, HiGHS) reproduces the certificates above (rep3 and rep8 with $N = 3$; rep1, rep2 and rep9 with $N = 4$; rep7 with $N = 5$; rep6 with $N = 6$; output `milp_designs_small.txt`) and finds none for rep5 and rep10 with $N = 6$ or $N = 7$, for every assignment of the columns to the classes with at least two columns of the pair class and one of the singleton class (for rep10 only designs with a copy of the rows). These restrictions lose nothing. For rep10 every admissible h' has cubic coefficient $-\frac{1+t}{8} \neq 0$, so every certificate carries a copy, on two columns of the pair class and one of the singleton class. For rep5, with $h' = h + t_1 \mathbf{1}_{+++} + t_2 \mathbf{1}_{-+-}$, $8\hat{h}'(\{1, 3\}) = t_1 + t_2 - 3$ and $8\hat{h}'(\{1, 2, 3\}) = t_1 + t_2 - 1$ do not both vanish, so both classes need a column. With a single column a of the pair class there is neither a pair of pair-class columns nor a copy, so $\hat{h}'(\{1, 2\}) = \hat{h}'(\{1, 2, 3\}) = 0$, which forces $t_1 = 1, t_2 = 0$; then $R = -\frac{1}{4} + \sum_j \beta_j b_j + a \sum_j \gamma_j b_j$ over the singleton-class columns b_j , with $\sum_j \beta_j = \hat{h}'(\{3\}) = \frac{1}{4}$ and $\sum_j |\gamma_j| \geq \frac{1}{2}$, since the coefficients of the pairs (a, b_j) sum to $\hat{h}'(\{1, 3\}) = -\frac{1}{4}$ on the law of (w_1, w_3) and to $\hat{h}'(\{2, 3\}) = \frac{1}{4}$ on the law of (w_2, w_3) , which is different. But $R \leq 0$ at $a = \pm 1$ requires $\sum_j |\beta_j \pm \gamma_j| \leq \frac{1}{4}$, hence $\sum_j |\gamma_j| \leq \frac{1}{4}$.

One-class designs. For a one-class h every column of a design carries the decoder F and has the law of X_1 , so any pair of columns may carry any pair law of π_b , and any triple a copy of the rows, with its columns in any order. This gives short proofs for two of the settled orbits.

(i) A24, the orbit of h_{g_1} : a third proof of the inequality $\Gamma_\nu \leq 0$ of Proposition 9.10, using neither Theorem K (as the proof of Proposition 9.10 does) nor the class flip from rep7 (Remark 11.11(b)). The representative is h_{g_1} itself (its $\pi_\pm$ are those of Proposition 9.10); in the order $+++ + + - , \dots, - - -$ of the vertices,

$$2h = (-2, -2, -1, 1, 0, 0, -1, -1), \quad \pi_+ = \frac{1}{10}(0, 0, 1, 3, 2, 2, 1, 1), \quad \pi_- = \frac{1}{22}(4, 4, 3, 1, 2, 2, 3, 3),$$

$\mu = -\frac{3}{8}$, and the three level-one coefficients equal $-\frac{1}{8}$ (one class). Let h' be h raised by 1 at $++-$, a vertex outside the support of π_+ . Then $h' - h = \mathbf{1}_{++-}$, whose multilinear extension $\frac{1}{8}(1 + y_1)(1 + y_2)(1 - y_3)$ is nonnegative on $[-1, 1]^3$, so $\Gamma_\nu(F) \leq \mathbb{E}[h'(Y)]$ with $Y = (F(X_1), F(X_2), F(X_3))$; $4h'(y) = -1 - y_3 - y_1 y_2 - y_1 y_3$, and, as the columns have a common law,

$$4\mathbb{E}[h'(Y)] = -1 - \mathbb{E}F(X_1) - \mathbb{E}[F(X_1)F(X_2)] - \mathbb{E}[F(X_1)F(X_3)].$$

Put

$$\begin{aligned} \tilde{\pi}_b(w) &= \pi_b(w) + c_b(1 - w_1)w_2w_3, & c_+ &= -\frac{1}{20}, & c_- &= \frac{1}{44}; \\ \tilde{\pi}_+ &= \frac{1}{10}(0, 0, 1, 3, 1, 3, 2, 0), & \tilde{\pi}_- &= \frac{1}{22}(4, 4, 3, 1, 3, 1, 2, 4), \end{aligned}$$

which are probability distributions. Summing $(1 - w_1)w_2w_3$ over w_3 for fixed (w_1, w_2) , or over w_2 for fixed (w_1, w_3) , gives 0, so $\tilde{\pi}_b$ has the same marginals as π_b on $\{1, 2\}$ and on $\{1, 3\}$; and under $\tilde{\pi}_b$ the pair (w_2, w_3) has the law of (w_1, w_3) under π_b (the values $++ , +- , -+ , --$ have probabilities $\frac{1}{10}(1, 3, 3, 3)$ for $b = +1$ and $\frac{1}{22}(7, 5, 5, 5)$ for $b = -1$). Let $\tilde{C}$ have independent rows, row $i \sim \tilde{\pi}_{\nu_i}$, and columns $\tilde{X}_1, \tilde{X}_2, \tilde{X}_3$. Then $(\tilde{X}_1, \tilde{X}_2)$ has the law of (X_1, X_2) , and $(\tilde{X}_1, \tilde{X}_3)$ and $(\tilde{X}_2, \tilde{X}_3)$ have that of (X_1, X_3) . Writing x_1, x_2, x_3 for $F(\tilde{X}_1), F(\tilde{X}_2), F(\tilde{X}_3)$ and splitting $\mathbb{E}F(X_1)$ and $\mathbb{E}[F(X_1)F(X_3)]$ evenly over two columns, respectively two pairs,

$$\begin{aligned} 4\mathbb{E}[h'(Y)] &= \mathbb{E}\left[-1 - \frac{1}{2}(x_1 + x_2) - x_1 x_2 - \frac{1}{2}(x_1 + x_2)x_3\right] \\ &= -\frac{1}{4}\mathbb{E}\left[(1 + x_1)(1 + x_2)(3 + x_3) + (1 - x_1)(1 - x_2)(1 - x_3)\right] \leq 0. \end{aligned}$$

In terms of $P = \frac{1+F}{2}$, $\Gamma_\nu(F) \leq -\frac{1}{2}\mathbb{E}[P(\tilde{X}_1)P(\tilde{X}_2)(1+P(\tilde{X}_3)) + (1-P(\tilde{X}_1))(1-P(\tilde{X}_2))(1-P(\tilde{X}_3))]$; both terms vanish for the dictator decoder on a row with $\nu_r = +1$, since $\tilde{\pi}_+(++\cdot) = \tilde{\pi}_+(- - -) = 0$. This is the device of Proposition 11.5 for a one-class h : h' has neither a cubic nor a y_2y_3 term, and the re-coupling changes only the law of (w_2, w_3) and the third moment, neither of which enters.

(ii) C12: a direct proof, without the class flip from rep6 of Proposition 11.12, and with $h' = h$. For the representative,

$$2h = (-1, -2, -2, 1, -1, 0, 0, -1), \quad \pi_+ = \frac{1}{10}(1, 0, 0, 3, 1, 2, 2, 1), \quad \pi_- = \frac{1}{22}(3, 4, 4, 1, 3, 2, 2, 3),$$

$\mu = -\frac{3}{8}$, one class, and $8h(y) = -3 - (y_1 + y_2 + y_3) - y_1y_2 - y_1y_3 + y_2y_3 + 3y_1y_2y_3$. As h is symmetric in w_2, w_3 , the pairs (X_1, X_2) and (X_1, X_3) have a common law, and

$$8\Gamma_\nu(F) = -3 - 3\mathbb{E}F(X_1) - 2\mathbb{E}[F(X_1)F(X_2)] + \mathbb{E}[F(X_2)F(X_3)] + 3\mathbb{E}[F(X_1)F(X_2)F(X_3)].$$

Let ρ_+ be the law on $\{-1, 1\}^4$ with $\rho_+(+--+)=\rho_+(-+--)=\rho_+(-+--)=\frac{1}{5}$ and $\rho_+(++++)=\rho_+(-+--)=\rho_+(-+--)=\frac{1}{10}$, and let $\rho_- = \frac{1}{11}(1 - 5\rho_+)$, which is 0 at the three atoms of mass $\frac{1}{5}, \frac{1}{22}$ at the four of mass $\frac{1}{10}$ and $\frac{1}{11}$ at the other nine points. Counting atoms, under ρ_+ the triples (u_1, u_2, u_3) and (u_4, u_2, u_3) have the law of (w_1, w_2, w_3) under π_+ , and (u_1, u_4) has the law of (w_2, w_3) ; the same holds for ρ_- and π_- by the complement argument of Proposition 11.8, as $(1+\mu)\rho_+ + (1-\mu)\rho_- = \frac{1}{8}$. Let U have independent rows, row $i \sim \rho_{\nu_i}$, and columns $U_1, \dots, U_4$, and write $a_j = F(U_j)$. Then (U_1, U_2, U_3) and (U_4, U_2, U_3) have the law of (X_1, X_2, X_3) ; so every U_j has the law of X_1 , the four pairs (U_t, U_u) with $t \in \{1, 4\}$, $u \in \{2, 3\}$ have the law of (X_1, X_2) , and (U_2, U_3) and (U_1, U_4) have that of (X_2, X_3) . Splitting the terms of $8\Gamma_\nu(F)$ over these columns, pairs and triples (the term $\mathbb{E}[F(X_2)F(X_3)]$ as $2\mathbb{E}[a_2a_3] - \mathbb{E}[a_1a_4]$),

$$\Gamma_\nu(F) = \mathbb{E}[R(a_1, a_2, a_3, a_4)],$$

$$16R = -6 - 3s - st - 2q + 4p + 3sp, \quad s = a_1 + a_4, \quad t = a_2 + a_3, \quad q = a_1a_4, \quad p = a_2a_3.$$

At the vertices, $16R = -14 - 2t + 10p$, which is $-8, -24, 0$ for $t = 2, 0, -2$, if $s = 2$; $16R = -4 + 4p \in \{0, -8\}$ if $s = 0$; and $16R = -2 + 2t - 2p$, which is $0, 0, -8$ for $t = 2, 0, -2$, if $s = -2$. So $R \leq 0$ on $[-1, 1]^4$, and $\Gamma_\nu(F) \leq 0$. As h has one class, conditions (c) and (d) of Theorem D coincide, so with Lemma 11.2 this re-proves Proposition 11.12 for the orbit C12. The polynomial R vanishes at the seven atoms of ρ_+ (dictator decoder).

The script `oneclass_certs.py` checks both certificates in exact arithmetic and tests the chains $\Gamma_\nu \leq \mathbb{E}[h'(Y)] = \mathbb{E}[R] \leq 0$ exactly on random decoders with $k \leq 3$. A search over one-class designs (`oneclass_designs.py`, floating point: pairs of columns carrying any pair law, triples carrying copies, all designs up to permutations of the columns) finds these two and a four-column design with pair laws only for A12 (the orbit of h_{g_0} ; not re-proved here), and none for B12 or D12 with $N \leq 5$; the mixed-integer formulation gives the same results (`milp_designs_small1.txt`). The mixed-integer formulation finds none for B12 with $N = 6$, and none for D12 with $N = 6$ that carries a copy of the rows. For B12 every design must carry a copy, since the only admissible h' is h raised at $+++$, whose cubic coefficient is at least $\frac{1}{8}$. Adding spectral inequalities gives no certificate for B12 with $N \leq 5$ either. For a pair law L of π_b let $\mathbb{E}_L$ refer to two columns x, y with independent rows, row i distributed as L under π_{ν_i} , and let $\lambda_L(b)$ be the correlation of L under π_b . As all columns have the law of X_1 , $\mathbb{E}_L[F(x)F(y)] = \sum_S \lambda_L^S \hat{F}(S)^2$ in the ν -biased basis, with $\lambda_L^S = \prod_{i \in S} \lambda_L(\nu_i)$; so if $\kappa_0 \geq 0$ and $\sum_L c_L \lambda_L^S \leq \kappa_0$ for all $S \neq \emptyset$, then, as $\mathbb{E}F^2 \leq 1$, $\sum_L c_L \mathbb{E}_L[F(x)F(y)] \leq (\sum_L c_L - \kappa_0)(\mathbb{E}F)^2 + \kappa_0$, and $(\mathbb{E}F)^2 = \mathbb{E}[F(V)F(V')]$ for two columns with independent law. (`milp_designs.py` imposes the conditions on S only for $|S| \leq 40$, which enlarges the feasible set, and bounds $|c_L| \leq 4$, $\kappa_0 \leq 1$,

which loses nothing here: $\kappa_0 \leq -\hat{h}'(\emptyset) \leq \frac{1}{4}$ and $|c_L| \leq |\hat{h}'_L| + 10|\mu| \leq 3$ for B12 with $N \leq 5$.) A design whose polynomial satisfies $R \leq \varepsilon$ instead of $R \leq 0$ proves $\Gamma_\nu(F) \leq \varepsilon$ for all k, ν and F , since $\Gamma_\nu(F) \leq \mathbb{E}[h'(Y)] = \mathbb{E}[R] \leq \max R$. For B12 the least such ε over designs with N columns is, in floating point (`eps_designs.py`), $\frac{1}{4}, \frac{1}{12}, \frac{1}{20}, \frac{1}{44}$ for $N = 3, 4, 5, 6$ (the optimal designs found have $h' = h$), and 0 for $N = 7$: minimising ε over designs with seven columns produced the certificate of Proposition 11.15. (Here $R \leq \varepsilon$ gives $|\hat{R}(S)| \leq \mathbb{E}[\varepsilon - R] = \varepsilon - \hat{h}'(\emptyset) \leq \varepsilon + |\mu|$ for $S \neq \emptyset$; as $\varepsilon \leq \frac{1}{4}$ is attained with $N = 3$, and a design extends to more columns by independent columns with the law of w_1 , the computations for $N = 4, 5, 6$ bound the coefficients by $\frac{1}{4} + |\mu| = \frac{1}{2}$, which loses nothing. For $N = 7$ the value 0 is exact without this: it is attained, and every design has $\max R \geq \mathbb{E}[R] \geq \Gamma_+ = 0$ for the dictator decoder.) Finally, Lemma 11.3 acts on designs: $h_{\text{B12}}(w) = h_{\text{rep10}}(w_2, w_1, -w_3)$ and $h_{\text{D12}}(w) = h_{\text{rep5}}(w_2, w_1, -w_3)$, and negating the singleton-class columns of a design for rep10 (rep5), together with the corresponding arguments of h' and R , and relabelling the coordinates by this permutation, gives a one-class design for B12 (D12) with the same number of columns: the marginal conditions transform as in the proof of Lemma 11.3, $R \leq 0$ is kept, and the two-class matching of the coefficients implies the one-class one. So a design for rep10 with N columns exists only if one exists for B12; in particular the computation for B12 with $N = 6$ independently excludes rep10 with $N = 6$.

Conjecture 11.14. For every h in `hard288.txt`, condition (d) of Theorem D fails, i.e. $s_g = \text{triv}_g$ for every g as in Theorem B with this h . Equivalently (Proposition 11.1): for $m = 5$, $s_g > \text{triv}_g$ iff $\text{sgn}(\mu)h(w) < 0$ for some w and $h \notin \text{hard288}$.

The conjecture is proved for 240 functions: the orbits of h_{g_0}, h_{g_1} and h_{g_2} (Theorem E, Propositions 9.10, 9.11; the first two again by Propositions 11.6 and 11.10, the second also by a three-column one-class design), the orbits rep3 and rep8 (Proposition 11.5), rep2 (Proposition 11.6), rep1 and rep9 (Proposition 11.8), rep7 (Proposition 11.10), rep6 and C12 (Proposition 11.12; C12 also by a four-column one-class design), and B12 (Proposition 11.15 below, a seven-column one-class design). It is open for the other 48: the two-class orbits rep5 and rep10 of size 24 (a proof for rep5 would give a second proof of Proposition 9.11, since rep5 implies D12). Neither has a certificate with $k \leq 3$ (Proposition 11.4), nor one with $k \leq 5$ whose pair-class decoder is $\{-1, 1\}$ -valued (exact computation, end of the paragraph “A reduction for rep10”; a decoder with fewer labels is one with $k = 5$ that ignores the extra labels), and a floating-point search finds no coupling design with at most seven columns for either (paragraph “Coupling designs”). The proofs of Theorem E and Propositions 9.10, 9.11 use three features. First, a single class, so that X, Y, Z have a common law and the pair expectations diagonalise in one basis. Second, a cubic term that can be removed by a pointwise bound with equality for the dictator decoder: $P_1 P_2 P_3 \leq P_1 P_2$ for g_0, g_1 (because $\pi_+(++\cdot) = 0$) and $P_1(1 - P_2)(1 - P_3) \leq P_1(1 - P_2)$ for g_2 (because $\pi_+(+-+) = 0$). In vertex terms, h is replaced by a multilinear $h' \geq h$ without cubic term that agrees with h on the support of π_+ . Third, in the resulting quadratic form only the coefficient of the level-one weight of the $\nu = +1$ rows is too large, and Theorem K (for g_0, g_1) or Theorem L (for g_2) controls that weight. For the one-class orbits B12 and C12 the second feature fails. For B12, $2h = (-2, -1, -1, 1, 0, 0, 0, -1)$, π_+ vanishes only at $+++$, and raising h there increases the cubic coefficient (which is positive), so no such h' exists. For C12, $2h = (-1, -2, -2, 1, -1, 0, 0, -1)$, all such h' give the same functional (by the symmetry $w_2 \leftrightarrow w_3$). That functional is positive for the $k = 1$ decoder with $P(1) = 1$, $P(-1) = \frac{1}{6}$, where it equals $\frac{1}{72}$, although $\Gamma_+ = -\frac{29}{144}$ there (`d12_check.py`); C12 is settled instead through rep6 (Proposition 11.12), and directly by a one-class design with $h' = h$ that keeps the cubic term (paragraph “One-class designs”). For B12 a floating-point search finds no design with at most six columns (same paragraph), but there is one with seven columns, again with $h' = h$; it was found by minimising ε (same paragraph).

Proposition 11.15. *Let g be as in Theorem B with h in the orbit B12 (Table 1; 12 functions). Then $s_g^L = s_g = \text{triv}_g < s_g^U$.*

Proof. For these h , $|\mu| = \frac{1}{4}$ and σh takes the value $-\frac{1}{2}$, where $\sigma = \text{sgn } \mu$, so $s_g^U > \text{triv}_g$ by Theorem C(ii). By Lemma 11.2 it suffices to show that $\Gamma_\nu(F) \leq 0$ for the representative h of B12 in Table 1, for all k, ν and $F: \{-1, 1\}^k \rightarrow [-1, 1]$ (there is one class, hence one decoder; as $\mu = -\frac{1}{4}$ for this h , this is the failure of condition (d) of Theorem D); then $s_g = \text{triv}_g$ by Theorem D, and $\text{triv}_g \leq s_g^L \leq s_g$.

In the order $+++ , ++- , \dots , ---$ of the vertices,

$$2h = (-2, -1, -1, 1, 0, 0, -1), \quad \pi_+ = \frac{1}{12}(0, 1, 1, 3, 2, 2, 1), \quad \pi_- = \frac{1}{20}(4, 3, 3, 1, 2, 2, 3),$$

$$h(y) = -\frac{1}{4} - \frac{1}{8}(y_1 + y_2 + y_3) - \frac{1}{4}(y_1 y_2 + y_1 y_3) + \frac{1}{8} y_1 y_2 y_3.$$

Both $\pi_\pm$ are invariant under exchanging w_2 and w_3 , and $\pi_+(w_j = 1) = \frac{5}{12}$ and $\pi_-(w_j = 1) = \frac{11}{20}$ for $j = 1, 2, 3$. So X_1, X_2, X_3 have a common law, so do (X_1, X_2) and (X_1, X_3) , and

$$8\Gamma_\nu(F) = -2 - 3\mathbb{E}F(X_1) - 4\mathbb{E}[F(X_1)F(X_2)] + \mathbb{E}[F(X_1)F(X_2)F(X_3)]. \quad (25)$$

The law of (w_1, w_2) is $\frac{1}{12}(1, 4, 4, 3)$ under π_+ and $\frac{1}{20}(7, 4, 4, 5)$ under π_- , on $++ , +- , -+ , --$. Write the points of $\{-1, 1\}^7$ as sign strings $u_0 u_1 \dots u_6$, and let ρ_+ (twelve points) and ρ_- (eighteen points) give the following masses to the points listed and 0 elsewhere:

$$\begin{array}{lcl} \rho_+ : \frac{1}{12} & ++----+- & +-+-----+ & +---+--+ & +----+--+ & +-----++ & -++-+- \\ & -++----- & -+-++++- & -+---++ & --+++- & --+++- & ----++ \\ \rho_- : \frac{1}{40} & ++++++- & ++++++- & +-+----- & +----+--+ & +-----+ & --+++- \\ & -----+ & & & & & \\ \frac{2}{40} & +++++++ & +++-++++ & +++-+- & +-+---+ & & \\ \frac{3}{40} & ++-++++ & +-+---+ & +-+---+ & --++++ & & \\ \frac{4}{40} & -++-+++ & -+-+--- & & & & \\ \frac{5}{40} & -----+ & & & & & \end{array}$$

A direct count shows, for $b = \pm 1$: (a) every coordinate u_j has the law of w_1 under π_b ; (b) each of the four triples (u_0, u_1, u_2) , (u_0, u_3, u_2) , (u_6, u_1, u_2) , (u_6, u_3, u_2) has the law π_b of (w_1, w_2, w_3) ; (c) each of the six pairs (u_1, u_3) , (u_0, u_4) , (u_2, u_4) , (u_4, u_5) , (u_2, u_5) , (u_5, u_6) has the law of (w_1, w_2) under π_b .

Let U have independent rows, row $i \sim \rho_{\nu_i}$, and columns $U_0, \dots, U_6$, and write $a_j = F(U_j)$. By (b) each of the four triples of columns listed there has the law of (X_1, X_2, X_3) ; so by (a)–(c) every U_j has the law of X_1 , and the four pairs (U_t, U_u) with $t \in \{0, 6\}$, $u \in \{1, 3\}$ and the six pairs of columns listed in (c) have the law of (X_1, X_2) . Splitting the terms of (25) over these columns, pairs and triples (the term $3\mathbb{E}F(X_1)$ equally over the six columns other than U_2 ; the term $4\mathbb{E}[F(X_1)F(X_2)]$ with weight $\frac{1}{4}$ on each of the four pairs (U_t, U_u) and $\frac{1}{2}$ on each of the six pairs of (c); the cubic term with weight $\frac{1}{4}$ on each of the four triples),

$$\begin{aligned} \Gamma_\nu(F) &= \mathbb{E}[R(a_0, \dots, a_6)], \\ 32R &= -8 - 2(a_0 + a_1 + a_3 + a_4 + a_5 + a_6) - (1 - a_2)(a_0 + a_6)(a_1 + a_3) \\ &\quad - 2(a_1 a_3 + a_0 a_4 + a_2 a_4 + a_4 a_5 + a_2 a_5 + a_5 a_6), \end{aligned}$$

where $(1 - a_2)(a_0 + a_6)(a_1 + a_3)$ collects the four pair terms of weight $\frac{1}{4}$ and the four cubic terms. As R is multilinear, it suffices to show $R \leq 0$ on $\{-1, 1\}^7$. Let $s = a_0 + a_6$ and $t = a_1 + a_3$, so that

$a_1a_3 = \frac{t^2}{2} - 1$, and let $\alpha = 1 + a_0 + a_2$, $\gamma = 1 + a_2 + a_6$. The terms of $32R$ that contain a_4 or a_5 are $-2(\alpha a_4 + a_4 a_5 + \gamma a_5)$, and the largest value of this over $a_4, a_5 \in \{-1, 1\}$ is $2 \max(|\alpha + \gamma| - 1, |\alpha - \gamma| + 1)$ (its values at $(a_4, a_5) = \pm(1, 1)$ are $\mp 2(\alpha + \gamma) - 2$, those at $\pm(1, -1)$ are $\mp 2(\alpha - \gamma) + 2$). With $\alpha + \gamma = 2 + 2a_2 + s$ and $\alpha - \gamma = a_0 - a_6$ this gives

$$32R \leq -6 - 2s - 2t - t^2 - (1 - a_2)st + 2 \max(|2 + 2a_2 + s| - 1, |a_0 - a_6| + 1).$$

If $a_2 = 1$, the maximum is $3 + s$ (note $|a_0 - a_6| = 2 - |s|$, so $32R \leq -t(t + 2)$). If $a_2 = -1$ and $s = 0$, the maximum is 3 and again $32R \leq -t(t + 2)$. If $a_2 = -1$ and $s = \pm 2$, the maximum is 1 and $32R \leq -4 - 2s - 2t - 2st - t^2$, which is $-(t + 2)(t + 4)$ for $s = 2$ and $t(2 - t)$ for $s = -2$. Since $t \in \{-2, 0, 2\}$, each of these bounds is at most 0. So $R \leq 0$ on $[-1, 1]^7$ and $\Gamma_\nu(F) \leq 0$. $\square$

The script `b12.cert.py` checks (a)–(c), the matching of the coefficients of R and h by law, the formula for R , the elimination of a_4, a_5 and $R \leq 0$ in exact arithmetic, and tests $\Gamma_\nu(F) = \mathbb{E}[R] \leq 0$ exactly on random decoders with $k \leq 3$ (all ν). R vanishes on the support of ρ_+ , as it must, since $\Gamma_\nu = 0$ for the dictator decoder. As in the proof of Proposition 11.12, the copies of the rows share a column (here U_2), and their cubic terms combine with pair terms; here into $-(1 - a_2)(a_0 + a_6)(a_1 + a_3)$.

Exact forms for rep5 and rep10. For a two-class h with pair class $\{1, 2\}$ and singleton class $\{3\}$, write $h(y) = c(y_1, y_2) + l(y_1, y_2)y_3$. For decoders F_p of the pair class and F_s of the singleton class, with $a_j = F_p(X_j)$, $\Gamma_\nu(F) = \mathbb{E}[c(a_1, a_2)] + \sum_z F_s(z) \mathbb{E}[l(a_1, a_2)\mathbf{1}\{X_3 = z\}]$ is affine in F_s , with one separate term for each value $F_s(z)$ (only the column X_3 carries F_s). So its maximum over F_s is attained at $F_s(z) = \text{sgn } \mathbb{E}[l(a_1, a_2) \mid X_3 = z]$ and equals $\mathbb{E}[c(a_1, a_2)] + \mathbb{E}|\mathbb{E}[l(a_1, a_2) \mid X_3]|$. For the representatives of rep5 and rep10 in Table 1,

$$\begin{aligned} h_{\text{rep5}}(y) &= -\frac{3}{8} - \frac{1}{8}(y_1 + y_2) + \frac{1}{8}y_3 - \frac{1}{8}y_1y_2 - \frac{3}{8}y_1y_3 + \frac{1}{8}y_2y_3 - \frac{1}{8}y_1y_2y_3, \\ h_{\text{rep10}}(y) &= -\frac{1}{4} - \frac{1}{8}(y_1 + y_2) + \frac{1}{8}y_3 - \frac{1}{4}y_1y_2 + \frac{1}{4}y_2y_3 - \frac{1}{8}y_1y_2y_3. \end{aligned}$$

Let $P = \frac{1+F_p}{2}$, $P_j = P(X_j)$ and $p = \mathbb{E}P_1$; then also $\mathbb{E}P_2 = p$, as X_1 and X_2 have the same law. Then $4l = 1 - 2P_1 + 2P_2 - 2P_1P_2$ and $4c = -1 - 2P_1P_2$ for rep5, and $4l = P_1 + 3P_2 - 2P_1P_2 - 1$ and $4c = P_1 + P_2 - 4P_1P_2 - 1$ for rep10. Writing $|x| = 2(-x)_+ + x$ for rep5 and $|x| = 2x_+ - x$ for rep10 gives

$$4 \max_{F_s} \Gamma_\nu(F) = 2 \left(\mathbb{E}[(\mathbb{E}[2P_1 - 1 - 2(1 - P_1)P_2 \mid X_3])_+] - 2\mathbb{E}[P_1P_2] \right) \quad (\text{rep5}),$$

$$4 \max_{F_s} \Gamma_\nu(F) = 2 \left(\mathbb{E}[(\mathbb{E}[P_1 + 3P_2 - 2P_1P_2 - 1 \mid X_3])_+] - p - \mathbb{E}[P_1P_2] \right) \quad (\text{rep10}).$$

So, by Lemma 11.2 (whose proof applies verbatim to class-wise decoders) and Theorem D, the 24 functions of rep5 are trivial iff (the part $\text{triv}_g < s_g^U$ of “trivial” holds for both orbits by Theorem C(ii), as $\sigma h(-, +, +) = -\frac{1}{2}$)

$$\mathbb{E}[(\mathbb{E}[2P_1 - 1 - 2(1 - P_1)P_2 \mid X_3])_+] \leq 2 \mathbb{E}[P_1P_2] \quad \text{for all } k, \nu \text{ and } P: \{-1, 1\}^k \rightarrow [0, 1], \quad (26)$$

and the 24 functions of rep10 are trivial iff

$$\mathbb{E}[(\mathbb{E}[P_1 + 3P_2 - 2P_1P_2 - 1 \mid X_3])_+] \leq \mathbb{E}P_1 + \mathbb{E}[P_1P_2] \quad \text{for all } k, \nu \text{ and } P: \{-1, 1\}^k \rightarrow [0, 1]. \quad (27)$$

Given the results above, Conjecture 11.14 is equivalent to (26) and (27). Both inequalities involve the joint law of the three columns, not only pair laws. Both are equalities for $P \equiv 0$ and for the

dictator $P(z) = \mathbf{1}\{z_r = 1\}$ with $\nu_r = +1$ (both sides are $\frac{1}{5}$ for rep5 and $\frac{1}{2}$ for rep10), (26) also for the AND of two labels with $\nu_r = \nu_s = +1$ (both sides $\frac{1}{50}$), and (27) also for the negated dictator (see below). The script `rep510_exact.py` checks the two identities exactly against a brute-force maximum of Γ_ν over all $\{-1, 1\}$ -valued F_s (random P , $k \leq 3$, every ν up to the order of the labels). Numerically both inequalities hold for $k \leq 8$: alternating maximisation of Γ_ν (exact one-dimensional steps in the values of F_p on $[-1, 1]$, best response in F_s ; constant, dictator, AND, $\mathbf{1}\{z_r = 1, z_s = -1\}$ and random starts, and the best decoders for $k - 1$ labels lifted by an ignored row) finds the maximum 0, up to $5 \cdot 10^{-14}$, for every $k \leq 8$ and every ν (`rep510_ascent.py`, floating point). As a positive control, the same search with the row laws of rep5 and the objective $h + \mathbf{1}_{+++}$, whose analogue of (28) fails (see below), finds the value 0.012 at $k = 3$.

For $k = 4$ both inequalities hold for every $\{0, 1\}$ -valued P and every ν , and among these P the equality cases are exactly $P \equiv 0$, the dictators and the ANDs of two labels for (26), and $P \equiv 0$, the dictators and the negated dictators $P(z) = \mathbf{1}\{z_r = -1\}$ for (27), where $\nu_r = \nu_s = +1$ throughout (exact integer arithmetic over all 2^{16} such P for each ν , `rep510_k4_boolean.py`). So $\Gamma_\nu \leq 0$ for every class-wise decoder with $k = 4$ whose pair-class decoder is $\{-1, 1\}$ -valued (Proposition 11.4 covers all $[-1, 1]$ -valued decoders, but only $k \leq 3$; for $k = 5$ see the end of the paragraph “A reduction for rep10”). For the negated dictator both sides of (27) are $\frac{5}{6}$, and the best response is $F_s \equiv 1$. On the slice $F_s \equiv 1$ the functional of rep10 is even in F_p : $h_{\text{rep10}}(-y_1, -y_2, 1) = h_{\text{rep10}}(y_2, y_1, 1) = -\frac{1}{8} + \frac{1}{8}(y_1 - y_2) - \frac{3}{8}y_1y_2$, and (X_1, X_2) is exchangeable (the law of (w_1, w_2) is $\frac{1}{12}(1, 4, 4, 3)$ under π_+ and $\frac{1}{20}(7, 4, 4, 5)$ under π_-), so $\Gamma_\nu(-F_p, 1) = \Gamma_\nu(F_p, 1)$ for every F_p ; in particular $\Gamma_\nu = 0$ at $(-F_p, 1)$ for the dictator F_p . Hence a coupling design that proves $\Gamma_\nu \leq 0$ for rep10 must have $R = 0$ also at the points obtained from the support of ρ_+ by negating the columns of the pair class and setting those of the singleton class to 1 (apply it to this decoder pair with $k = 1$, $\nu = (+1)$).

A reduction for rep5. Let $2h = (-2, -1, -1, 0, 1, -2, 0, -1)$ be the representative of rep5 in Table 1. Its pair class is $\{1, 2\}$, its singleton class is $\{3\}$, $\mu = -\frac{3}{8}$, and π_+ vanishes exactly at $+++$ and $-+-$. The function $h' = h + \mathbf{1}_{-+-}$ agrees with h on the support of π_+ , satisfies $h' \geq h$ on $[-1, 1]^3$ (the multilinear extension of $\mathbf{1}_{-+-}$ is $\frac{1}{8}(1 - y_1)(1 + y_2)(1 - y_3) \geq 0$), and has no cubic term: $4h'(y) = -1 - y_1(1 + y_2 + y_3)$. For decoders F_p of the pair class and F_s of the singleton class, and $Y = (F_p(X_1), F_p(X_2), F_s(X_3))$,

$$4\Gamma_\nu(F) \leq 4\mathbb{E}[h'(Y)] = -1 - \mathbb{E}F_p(X_1) - \mathbb{E}[F_p(X_1)F_p(X_2)] - \mathbb{E}[F_p(X_1)F_s(X_3)].$$

The right-hand side is affine in F_s and largest for $F_s(z) = -\text{sgn } \mathbb{E}[F_p(X_1) \mid X_3 = z]$, where, in terms of $P = \frac{1+F_p}{2}$, it equals $2\mathbb{E}[(2\mathbb{E}[P(X_1) \mid X_3] - 1)_+] - 4\mathbb{E}[P(X_1)P(X_2)]$. So, by Lemma 11.2 (whose proof applies verbatim to class-wise decoders), Theorem D and Theorem C(ii) ($\sigma h(-, +, +) = -\frac{1}{2}$), $s_g^L = s_g = \text{triv}_g < s_g^U$ for the 24 functions of rep5 if

$$\mathbb{E}\left[(2\mathbb{E}[P(X_1) \mid X_3] - 1)_+\right] \leq 2\mathbb{E}[P(X_1)P(X_2)] \quad \text{for all } k, \nu \text{ and } P: \{-1, 1\}^k \rightarrow [0, 1], \quad (28)$$

an inequality of the shape of (30). It strengthens (26) by dropping the term $-2(1 - P_1)P_2$ inside the conditional expectation, which removes the dependence on the joint law of the three columns. Only two pair laws enter: on $++$, $+-$, $-+$, $--$, the law of (w_1, w_2) is $\frac{1}{10}(1, 3, 3, 3)$ under π_+ and $\frac{1}{22}(7, 5, 5, 5)$ under π_- , and that of $(w_1, -w_3)$ is $\frac{1}{10}(3, 1, 1, 5)$ and $\frac{1}{22}(5, 7, 7, 3)$. These are also the laws of (w_1, w_2) and (w_2, w_3) in the proof of Proposition 9.11. For $P, B: \{-1, 1\}^k \rightarrow [0, 1]$ let $\Psi(P, B) = 2\mathbb{E}[P(X_1)P(X_2)] - 2\mathbb{E}[P(X_1)B(-X_3)] + \mathbb{E}[B(-X_3)]$. As Ψ is affine in B , its minimum over B is attained at $B(v) = \mathbf{1}\{2\mathbb{E}[P(X_1) \mid X_3 = -v] > 1\}$ and equals the right side of (28)

minus the left side; so (28) says $\Psi \geq 0$. For $B = P$, $\Psi(P, P) \geq 0$ is the inequality $-p - 2\mathbb{E}[P_1 P_2] + 2\mathbb{E}[P_2 P_3] \leq 0$ established in the proof of Proposition 9.11 (in that proof's notation: $\mathbb{E}[P(X_1)P(X_2)]$, $\mathbb{E}[P(X_1)P(-X_3)]$ and $\mathbb{E}P(-X_3)$ are its $\mathbb{E}[P_1 P_2]$, $\mathbb{E}[P_2 P_3]$ and p ; here $\mathbb{E}P(-X_3) = \mathbb{E}P(X_1)$, as $-w_3$ and w_1 have the same law under $\pi_\pm$, with probabilities $\frac{2}{5}$ and $\frac{6}{11}$ of $+1$), so (28) is a two-function form of it (in accordance with “rep5 implies D12”). That proof uses $B = P$: the cross term is then a diagonal quadratic form in the Fourier coefficients of P .

Inequality (28) is an equality for $P \equiv 0$, for the dictator $P(z) = \mathbf{1}\{z_r = 1\}$, for the AND of two labels $P(z) = \mathbf{1}\{z_r = z_s = 1\}$ and for $P(z) = \mathbf{1}\{z_r = 1, z_s = -1\}$, with $\nu_r = \nu_s = +1$ (both sides are $0, \frac{1}{5}, \frac{1}{50}$ and $\frac{3}{50}$). Among the $\{0, 1\}$ -valued P that depend on all of $k \leq 3$ labels there are no other equality cases, and (28) holds for all of them (exact enumeration over all ν , `rep5_reduction.py`). It also holds for every $[0, 1]$ -valued P with $k \leq 3$, and for every $\{0, 1\}$ -valued P with $k = 4$ or $k = 5$, with no equality cases among the latter other than these four kinds (exact; see the paragraph “A reduction for rep10”). The dictator and the AND are the two decoders at which every inequality in the proof of Proposition 9.11 is an equality. The third one also has $\Psi(P, P) = 0$: every inequality in that proof after the first is an equality for it (it is an equality case of Theorem L, and its only coefficient above level one is at $\{r, s\}$). But for rep5 at it and its best response $\Gamma_\nu = -\frac{1}{100}$: the vertex $-+-$, where $h' - h = 1$, then has probability $\frac{1}{100}$. Correspondingly $\Gamma_\nu = 0$ for rep5 at the dictator pair and at the pair with $F_p(z) = 1$ iff $z_r = z_s = 1$ and $F_s(z) = -1$ iff $z_r = z_s = -1$ ($\nu_r = \nu_s = +1$). Hence a coupling design that proves $\Gamma_\nu \leq 0$ for rep5 must have $R = 0$ not only on the support of ρ_+ but also at every vertex obtained from two points of that support by taking the coordinatewise minimum on the columns of the pair class and the maximum on those of the singleton class (apply the design to this decoder pair with $k = 2$, $\nu = (+1, +1)$: then $0 = \Gamma_\nu \leq \mathbb{E}[h'(Y)] = \mathbb{E}[R] \leq 0$). The same holds for D12, with the minimum on every column. Numerically, (28) holds: alternating maximisation of its left side minus its right side (best response in B , exact coordinate steps in P , in each of which the objective is concave; dictator, constant, AND and OR, and random starts) gives the maximum 0, up to $2 \cdot 10^{-16}$, for every $k \leq 8$ and every ν (`rep5_relax.py`). The choice of h' matters: the cubic-free $h' \geq h$ that agree with h on the support of π_+ are $h + t_1 \mathbf{1}_{++++} + t_2 \mathbf{1}_{-+-}$ with $t_1 + t_2 = 1$, and the analogue of (28) also holds numerically for $t_1 = \frac{1}{4}$ and $t_1 = \frac{1}{2}$ ($k \leq 4$), but fails for $t_1 = 1$ ($k = 3$, $\nu = (+1, +1, +1)$), where the exact maximum of $\mathbb{E}[(h + \mathbf{1}_{++++})(Y)]$ over all $[-1, 1]$ -valued class-wise decoders is $\frac{3}{250}$, attained by $\{-1, 1\}$ -valued ones; `majorant_k3_exact.py`). On the side of coupling designs, the least ε of the paragraph “One-class designs” over designs with N columns (for two classes, with at least two columns of the pair class and one of the singleton class; floating point, solver status optimal, coefficients bounded by $\frac{1}{2} + |\mu|$, which loses nothing as the three-column design with $\rho_b = \pi_b$ and $R = h'$ already gives $\varepsilon = \max h' = \frac{1}{2}$) is $\frac{1}{4}, \frac{1}{6}, \frac{1}{10}, \frac{1}{14}, \frac{1}{20}$ for rep5 and $\frac{1}{4}, \frac{1}{6}, \frac{1}{12}, \frac{1}{18}, \frac{1}{24}$ for rep10, for $N = 3, \dots, 7$, against $\frac{1}{4}, \frac{1}{12}, \frac{1}{20}, \frac{1}{44}, 0$ for B12. For D12, which is trivial by Proposition 9.11, it is $\frac{1}{4}, \frac{1}{12}, \frac{1}{18}, \frac{1}{30}$ for $N = 3, \dots, 6$ (`eps_designs.py`). So the absence of short designs for rep5 and rep10 says little about Conjecture 11.14. On the other hand, since a design with $\max R \leq \varepsilon$ proves $\Gamma_\nu \leq \varepsilon$ for all k, ν and F , the conjecture for rep5 (rep10) would follow if the infimum of ε over designs with N columns tended to 0 as $N \rightarrow \infty$, even if no single design gives 0. The values above decrease with N ; for rep10 they are $\frac{1}{6(N-3)}$ for $4 \leq N \leq 7$, and the interrupted searches with $N = 8$ found designs with $\varepsilon \leq \frac{1}{33}$ (floating-point incumbents; a run of 40 minutes for the pattern with six pair-class columns, restricted to $\varepsilon \leq \frac{1}{24}$ with the corresponding cap $\frac{1}{24} + |\mu|$ on the coefficients, also stopped at $\frac{1}{33}$). Two-factor identities of the kind behind the proof for D12 do not suffice either. The SDP search described below after (30) is infeasible for rep5 as well (output `cert_sdp_rep5.txt`). The level-weight relaxation of Ψ is tight at the three nonzero equality cases, but its minimum is negative, about -0.015 (`rep5_scalar.py`, floating point). It bounds the level-one parts on the rows with $\nu_i = +1$ by Theorem L (for P, B and $P + cB$, $c > 0$)

and Cauchy–Schwarz, and every other level by the arithmetic–geometric mean inequality. As for rep3, the minimum is at small means ($p \approx 0.1$, $\mathbb{E}B \approx 0.025$), where Theorem L is far from sharp.

A reduction for rep10. The device of the preceding paragraph applies to rep10, with a single free vertex. A majorant can only work if it agrees with h on the support of π_+ , since the dictator decoders with $\nu_r = +1$ have $Y \sim \pi_+$ and $\Gamma_\nu = 0$. For $h = h_{\text{rep10}}$ the law π_+ vanishes only at $++-$, where $h = -1$. So a multilinear $h' \geq h$ on $[-1, 1]^3$ that agrees with h on the support of π_+ is $h + t\mathbf{1}_{++-}$ with $t \geq 0$ (a multilinear function is nonnegative on $[-1, 1]^3$ iff it is nonnegative at the vertices), and $\mathbb{E}[h'(Y)] \leq 0$ for all decoders forces $t \leq 1$, as the constant decoders $F_p \equiv 1$, $F_s \equiv -1$ give $\mathbb{E}[h'(Y)] = t - 1$. The largest one, $h'' = h + \mathbf{1}_{++-}$, has no linear terms,

$$8h''(y) = -1 - y_1y_2 - y_1y_3 + y_2y_3 - 2y_1y_2y_3,$$

and satisfies $h''(-y_1, -y_2, y_3) = h''(y_2, y_1, y_3)$, which extends the identity on the slice $y_3 = 1$ used above (on that slice $h'' = h$). In the notation of the paragraph “Exact forms for rep5 and rep10”, $4c'' = P_1 + P_2 - 2P_1P_2 - 1$ and $4l'' = P_1 + 3P_2 - 4P_1P_2 - 1$, and the same computation gives $4\max_{F_s} \mathbb{E}[h''(Y)] = 2(\mathbb{E}[(\mathbb{E}[4l'' \mid X_3])_+] - \mathbb{E}[P_1(1 - P_2)])$. So, as $\Gamma_\nu(F) \leq \mathbb{E}[h''(Y)]$, the 24 functions of rep10 are trivial if

$$\mathbb{E}\left[(\mathbb{E}[P_1 + 3P_2 - 4P_1P_2 - 1 \mid X_3])_+\right] \leq \mathbb{E}[P_1(1 - P_2)] \quad \text{for all } k, \nu \text{ and } P: \{-1, 1\}^k \rightarrow [0, 1], \quad (29)$$

which implies (27). The set of decoder pairs is invariant under $(F_p, F_s) \mapsto (-F_p, -F_s)$, which fixes the quadratic part of h'' and negates its cubic term. So (29) is equivalent to

$$2|\mathbb{E}[a_1a_2b_3]| \leq 1 + \mathbb{E}[a_1a_2] + \mathbb{E}[a_1b_3] - \mathbb{E}[a_2b_3] \quad (a_j = F_p(X_j), b_3 = F_s(X_3))$$

for all k, ν and $[-1, 1]$ -valued F_p, F_s : a bound on one cubic correlation of the three columns by pair correlations. With $Q = 1 - P$ and $B = \frac{1+F_s}{2}$, the identity $-8h'' = 4(\mathbf{1}_{+++} + \mathbf{1}_{-+-} + \mathbf{1}_{--+} - \mathbf{1}_{-++})$ gives a third form,

$$4\mathbb{E}[Q_1P_2B_3] \leq \mathbb{E}[Q_1P_2] + \mathbb{E}[Q_1B_3] + \mathbb{E}[P_2B_3] \quad (Q_1 = Q(X_1), B_3 = B(X_3))$$

for all $P, B: \{-1, 1\}^k \rightarrow [0, 1]$. If the coordinates of Y are rounded independently to ± 1 with conditional means Y_j , it says that the vertex $-++$ is at most as likely as its three neighbours $+++$, $--+$ and $-+-$. With 3 in place of 4 this holds for any three events, and for the dictator ($Y \sim \pi_+$) both probabilities are $\frac{1}{4}$. For $k = 4$, (29) holds for every $\{0, 1\}$ -valued P and every ν , and among these P the equality cases are exactly $P \equiv 0$, $P \equiv 1$, and the dictators and negated dictators on labels with $\nu_r = +1$ (exact integer arithmetic over all 2^{16} such P for each ν , `rep10_majorant.py`; the script also checks exactly the statements above about π_+ and h'' , and the one-function form against a brute-force maximum over F_s for random P with $k \leq 3$). The constant $P \equiv 1$ is an equality case of (29) only: for h itself, $\max_{F_s} \Gamma_\nu = h(+++) = -\frac{1}{2}$ there. For $k \leq 3$, (29) holds for every $[0, 1]$ -valued P and every ν : the part of the proof of Proposition 11.4 after the reduction to the representatives uses only that the integrand is a multilinear function of $(F_p(X_1), F_p(X_2), F_s(X_3))$, so that $\mathbb{E}[h''(Y)]$ is affine in the values of F_s and quadratic in those of F_p ; hence its face enumeration applies verbatim to $\mathbb{E}[h''(Y)]$ (with the row laws $\pi_\pm$ of h), and the exact maximum of $\mathbb{E}[h''(Y)]$ over all $[-1, 1]$ -valued class-wise decoders with $k \leq 3$ is 0 for every ν (`majorant_k3_exact.py`, exact rational arithmetic). The same computation for the majorant $h + \mathbf{1}_{-+-}$ of rep5 gives the maximum 0, which is (28) for every $[0, 1]$ -valued P with $k \leq 3$, and for $h + \mathbf{1}_{+++}$ the positive maximum $\frac{3}{250}$ quoted in the previous paragraph. The script also checks (28)

for all 2^{16} $\{0, 1\}$ -valued P with $k = 4$ and every ν , in integer arithmetic; its equality cases there are $P \equiv 0$, the dictators, the ANDs of two labels and $\mathbf{1}\{z_r = 1, z_s = -1\}$, all on labels with $\nu = +1$. Beyond $\{0, 1\}$ -valued P at $k = 4$: all four inequalities (26), (27), (28) and (29) hold for every P with values in $\{0, \frac{1}{2}, 1\}$ (all 3^{16} of them, for each ν), and none of these P that takes the value $\frac{1}{2}$ is an equality case (`rep510_k4_half.py`, exact integer arithmetic). For $k = 5$ the four inequalities hold for every $\{0, 1\}$ -valued P (all 2^{32} of them, for each ν), and the equality cases are exactly the families found for $k = 4$, on labels with $\nu = +1$. These are $P \equiv 0$, the dictators and the ANDs of two labels for (26); these and $\mathbf{1}\{z_r = 1, z_s = -1\}$ for (28); $P \equiv 0$, the dictators and the negated dictators for (27); and these and $P \equiv 1$ for (29) (`rep510_boolean.py`, exact integer arithmetic). In particular $\Gamma_\nu \leq 0$ for rep5 and rep10 for every class-wise decoder with $k = 5$ whose pair-class decoder is $\{-1, 1\}$ -valued. In all these computations it suffices to take $\nu = (+1)^r(-1)^{k-r}$, $0 \leq r \leq k$, and the scripts do so: a permutation of the labels permutes the independent rows of X together with their laws $\pi_{\nu_1}, \dots, \pi_{\nu_k}$, acts on P (and on the decoders F_p, F_s) by the same permutation of coordinates (which preserves the value sets $\{0, 1\}$, $\{0, \frac{1}{2}, 1\}$, $[0, 1]$ and $[-1, 1]$), leaves Γ_ν and both sides of each of the four inequalities unchanged, and maps each family of equality cases above to the corresponding family on the permuted labels. Numerically (29) holds for every $k \leq 8$ and every ν : the alternating maximisation of the paragraph “Exact forms for rep5 and rep10”, with the objective h'' in place of h , finds the maximum 0, up to $7 \cdot 10^{-15}$ (`rep510_ascent.py`, floating point; 200 random starts for each ν with $k \leq 5$). For rep5, by contrast, the majorant $h + \mathbf{1}_{+++}$, which raises h at $+++$ by the largest amount that the constant decoders allow, already fails (previous paragraph).

Numerical evidence for Conjecture 11.14. Coordinate ascent (exact one-dimensional steps, floating point) over single $[-1, 1]$ -valued decoders F on $\{-1, 1\}^k$ reaches $k = 6$ for the eleven orbits B12, C12 and the nine two-class orbits of Table 1 (all except rep5 and rep10 have since been settled, by Propositions 11.5–11.12 and 11.15) and $k = 5$ for the settled orbits A12 and A24 (D12 was not run); each step maximises the cubic $t \mapsto \Gamma_\nu(F)$ in one value $F(z) = t$ exactly, from dictator-type and random starts, for every ν up to permutation (script `orbit_coord_ascent.py`). Class-wise ascent with one decoder per class reaches $k = 5$ for all nine two-class orbits (script `classwise_ascent.py`). No run finds a value of $\sigma\Gamma_\nu$ below $-2 \cdot 10^{-15}$: the largest value of Γ found is 0 up to rounding. For the two-class orbits, Γ_ν is linear in the decoder of the singleton class, so the best response to the other decoder is a sign pattern, and the problem becomes an L^1 inequality of the type of Theorem K for the pair class. For instance, for rep3 (pair class $\{1, 3\}$ with decoder $P = \frac{1+F}{2}$, singleton class $\{2\}$) one finds $2\Gamma_\nu = \mathbb{E}[P_2(2P_3 - 2P_1 - 1)] - \mathbb{E}[P_1P_3]$, where $P_j = P(C_{\cdot j})$ for $j = 1, 3$ and $P_2 = P'(C_2)$ for the decoder P' of the singleton class (using $\mathbb{E}P_1 = \mathbb{E}P_3$). Taking P' optimal, triviality of rep3 is the inequality

$$\mathbb{E}\left[\left(\mathbb{E}[2P(C_{\cdot 3}) - 2P(C_{\cdot 1}) - 1 \mid C_{\cdot 2}]\right)_+\right] \leq \mathbb{E}[P(C_{\cdot 1})P(C_{\cdot 3})] \quad \text{for all } k, \nu, P,$$

with equality for the dictator (both sides are $\frac{1}{10}$). Since $P \geq 0$, removing the term $-2P(C_{\cdot 1})$ can only increase the left side, which gives the stronger inequality

$$\mathbb{E}\left[\left(2\mathbb{E}[P(C_{\cdot 3}) \mid C_{\cdot 2}] - 1\right)_+\right] \leq \mathbb{E}[P(C_{\cdot 1})P(C_{\cdot 3})] \quad \text{for all } k, \nu \text{ and } P: \{-1, 1\}^k \rightarrow [0, 1]. \quad (30)$$

In vertex terms this replaces h by the function h' of the proof of Proposition 11.5, and (30) is the statement $\mathbb{E}[h'(Y)] \leq 0$ of that proof for the best response P' ; so (30) holds. Only the pair laws of $(C_{\cdot 2}, C_{\cdot 3})$ and of $(C_{\cdot 1}, C_{\cdot 3})$ enter, and (30) is an equality for $P \equiv 0$, for $P \equiv 1$ and for $P(z) = \mathbf{1}\{z_r = 1\}$ and $P(z) = \mathbf{1}\{z_r = -1\}$ when $\nu_r = +1$ (both sides are then $\frac{1}{10}$, respectively $\frac{3}{10}$). Before the proof was found, alternating maximisation of the left side minus the right side

(best response in the decoder of C_2 , exact coordinate steps in P , from dictator, constant and random starts; `rep3_star.py`) gave the maximum 0 for every $k \leq 7$ and every ν , and the level-weight method of Theorem E and Propositions 9.10 and 9.11 was found not to suffice: an SDP search (`cert_sdp.py`) for an identity that writes $-2\Gamma'_\nu$, for some cubic-free $h' \geq h$ equal to h on the support of π_+ , as a nonnegative combination of terms $\mathbb{E}[A(x)B(y)]$ with $A, B \in \{P, 1-P, P', 1-P'\}$ and (x, y) a product coupling (five per-row correlations for each row sign, spread over the admissible range), plus a quadratic polynomial in $(\mathbb{E}P, \mathbb{E}P')$ written as a sum of squares plus a nonnegative combination of the products of the constraints $0 \leq \mathbb{E}P, \mathbb{E}P' \leq 1$, plus positive semidefinite 2×2 forms in $(\hat{P}(S), \hat{P}'(S))$, $|S| \leq 8$, reports that no such identity exists (floating point). The certificate of Proposition 11.5 lies outside this search space: after re-coupling the rows it writes the (quadratic) integrand as a sum of nonnegative products of three factors from $\{P, 1-P, P', 1-P'\}$, evaluated at three columns of $\tilde{C}$, whereas the terms above have two factors.

The dictator is a strict local maximum. For each representative the dictator decoder ($F_c(z) = z_1$ for every class, $\nu_1 = +1$) has $\Gamma_\nu = 0$, and at it $\partial\Gamma_\nu/\partial F_c(z) = D_c(z_1) \prod_{i \geq 2} \pi_{\nu_i}(w_j = z_i)$ for any $j \in c$: the partial derivatives of h at the dictator depend on row 1 only, the rows are independent, and $\pi_b(w_j = s)$ depends on j only through its class. Here $D_c(s) = \sum_w \pi_+(w) \sum_{j \in c, w_j = s} \partial_j h(w)$ is the derivative for $k = 1$. For all representatives except `rep10`, $D_c(1) > 0 > D_c(-1)$ for every class and all marginals $\pi_b(w_j = s)$ are positive (`orbit_kkt.py`, exact). So every admissible direction (lowering $F_c(z)$ where it is 1, raising it where it is -1) decreases Γ_ν at first order: the dictator is a strict local maximum of Γ_ν for every k and every ν with $\nu_1 = +1$, and a certificate would have to be far from it. For `rep10` the derivative of the singleton class $\{3\}$ at $s = -1$ vanishes. As Γ_ν is affine in the decoder of a singleton class, $\Gamma_\nu = 0$ on the whole face where $F_{\{3\}}(z)$, $z_1 = -1$, varies in $[-1, 1]$ (for every k). On this face the first-order conditions remain strict in all other directions, for every k . The derivative in $F_{\{3\}}(z)$, $z_1 = 1$, does not depend on the point of the face and equals $D_{\{3\}}(1) \prod_{i \geq 2} \pi_{\nu_i}(w_3 = z_i) > 0$, with $D_{\{3\}}(1) = \frac{1}{8}$. For the pair class,

$$\frac{\partial\Gamma_\nu}{\partial F_{\{1,2\}}(z)} = A(z_1) \prod_{i \geq 2} m_{\nu_i}(z_i) + \sum_{j=1,2} \alpha_j(z_1) \sum_{z' \in \{-1,1\}^{k-1}} F_{\{3\}}(-1, z') \prod_{i \geq 2} \pi_{\nu_i}(w_j = z_i, w_3 = z'_i),$$

where $m_b(s) = \pi_b(w_1 = s) = \pi_b(w_2 = s)$, $A(s)$ is the derivative for $k = 1$ at the face point $F_{\{3\}}(-1) = 0$, and $\alpha_j(s) = \sum_{w: w_j = s, w_3 = -1} \pi_+(w) \partial_j \partial_3 h(w)$. Since $|F_{\{3\}}| \leq 1$ and summing the pair laws over z'_i gives the marginals, $s \partial\Gamma_\nu/\partial F_{\{1,2\}}(z) \geq (sA(s) - |\alpha_1(s)| - |\alpha_2(s)|) \prod_{i \geq 2} m_{\nu_i}(z_i)$ for $s = z_1$, and the bracket equals $\frac{13}{96} - \frac{1}{48} - \frac{1}{32} = \frac{1}{12}$ for $s = 1$ and $\frac{5}{32} - \frac{1}{96} - \frac{1}{12} = \frac{1}{16}$ for $s = -1$ (`orbit_kkt.py`, exact). As the face is compact and Γ_ν is a polynomial, it follows that $\Gamma_\nu \leq 0$ on a neighbourhood of the face, with equality only on the face, for every k and every ν with $\nu_1 = +1$. So for all 14 orbits a certificate would have to be far from the dictator decoder.

Two relevant coordinates, nine variables. For $r = 7$ ($m = 9$) the computation of Proposition 10.7 (`maxcorr_J2.py` 7) gives TRIV 133120, CERT 374206, F 15870, F^+ 828, $F^\#$ 208 and CERT' 20 (8 with a certificate with $k = 2$ and 12 with $k = 3$), and leaves 34 candidates undecided, forming 17 pairs $\{h, h(-\cdot)\}$. For each of them (`m9_open.py`, exact) $\varepsilon = 1$ and exactly one sign is dangerous, $\Phi^\#$ is negative exactly on a subinterval of (q^2, q) (of $(1-q, 1-q^2)$ for the mirror images), its minimum lies between $-2.6 \cdot 10^{-3}$ and $-4.7 \cdot 10^{-5}$, and the maximum of Γ_ν over $[-1, 1]$ -valued decoders with $k \leq 3$ is negative. This extends to four labels:

Proposition 11.16 (computer-assisted). *Let h be one of the 34 candidates left open for $m = 9$. Then $\Gamma_\nu(F) < 0$ for every $k \leq 4$, every $\nu \in \{-1, 1\}^k$ and every decoder $F: \{-1, 1\}^k \rightarrow [-1, 1]$. So none of them has a certificate as in Theorem D(c) with at most four label bits.*

Proof. All candidates have $\mu < 0$, so we must show $\Gamma_\nu(F) < 0$. A decoder with $k < 4$ labels is a decoder with four labels that ignores the last ones, with the same value of Γ for any extension of ν (the additional rows of C are independent of the others), so it suffices to take $k = 4$. The map $F \mapsto \tilde{F}$, $\tilde{F}(z) = -F(-z)$, in the proof of Lemma 11.2 keeps k and ν , so it suffices to treat one candidate of each pair $\{h, h(-\cdot)\}$, and permuting the rows of C (and the label bits) lets us take $\nu = (1, \dots, 1, -1, \dots, -1)$. Fix h and such a ν , and multiply Γ_ν by the common denominator of its coefficients: it becomes a quadratic polynomial in the 16 values of F with integer coefficients, the sum of whose absolute values (with the entries of the symmetric matrix of the quadratic part counted twice) is below 2^{52} . As in the proof of Proposition 10.9, its maximum over $[-1, 1]^{16}$ is attained at a vertex or at the critical point of its restriction to the relative interior of a face on which the Hessian block H is nonsingular. For a face with free set T and fixed values $a \in \{-1, 1\}^{T^c}$ the restriction is $\beta(a) + \langle b(a), s \rangle + \frac{1}{2} \langle s, Hs \rangle$, $s \in [-1, 1]^T$, where $\beta(a)$ and the vector $b(a)$ are integers that are computed exactly in double precision (every partial sum is a signed sum of some of the integer coefficients, so its modulus is below 2^{52} whatever the order of summation); at the critical point $S = -H^{-1}b(a)$ the value is $\beta(a) + \frac{1}{2} \langle b(a), S \rangle$. For each T the script `m9_k4_rig.py` computes an approximate inverse R of H in floating point and a rigorous upper bound α for $\|I - RH\|_\infty$, in which the rounding errors of the products are bounded by $\gamma_t |R| |H|$ entrywise, $\gamma_t = \frac{tu}{1-tu}$, $u = 2^{-53}$, $t = |T|$. (The script uses $2\gamma_t$ and multiplies α , $\|R\|_\infty$ and the enclosure below by $1 + 10^{-10}$; the value bound carries an extra $10^{-12}(|\beta(a)| + |\langle b(a), \hat{S} \rangle| + \langle |b(a)|, e \rangle + \langle |b(a)|, \hat{S} \rangle + 1)$, where $\hat{S}$ is the computed point and e its error bound. These margins absorb the rounding errors made in evaluating the bounds themselves, which are of relative size below 10^{-14} .) If $\alpha < \frac{1}{2}$, then H is nonsingular and $\|H^{-1} - R\|_\infty \leq \frac{\alpha}{1-\alpha} \|R\|_\infty \leq 2\alpha \|R\|_\infty$, so every coordinate of S differs from that of the computed $-Rb(a)$ by at most $2\alpha \|R\|_\infty \|b(a)\|_\infty + \gamma_t (|R| |b(a)|)_i$. Critical points whose enclosure meets no point of $[-1, 1]^T$ are discarded; for the others an upper bound for the value is computed, again with all rounding errors bounded. Blocks with $\alpha \geq \frac{1}{2}$, and critical points whose upper bound is not negative, are treated in exact rational arithmetic (python-flint): a block with $\det H = 0$ is skipped, and otherwise the critical point and its value are computed exactly. Vertices are evaluated exactly. In the run every block has $\alpha < \frac{1}{2}$, every upper bound is negative, and the largest of them, over all faces, vertices, ν and pairs, is $-7.17 \cdot 10^{-5}$ (output `m9_k4_rig.txt`). The floating-point computations assume IEEE double precision with rounding to nearest and classical matrix products, for which the bounds γ_t hold for every order of summation, with or without fused multiply-add. Underflow does no harm: in every matrix and vector product one factor has integer entries (H or $b(a)$), so a product that falls in the subnormal range is computed exactly, as is every sum, and the remaining quantities (α , $\|R\|_\infty$, the margins) stay far above that range. $\square$

The bound $-7.17 \cdot 10^{-5}$ is attained up to the margins: for $h = (-\frac{23}{32}, \frac{5}{64}, \frac{5}{64}, -\frac{1}{32})$ and $\nu = (1, 1)$ the decoder $F(z) = 1$ if $(z_1, z_2) = (-1, 1)$ and $F(z) = -1$ otherwise has $\Gamma_\nu(F) = -\frac{1}{13952} = -7.16743 \dots \cdot 10^{-5}$, which is also the exact maximum over decoders with $k \leq 3$ (`m9_open.py`). Two checks of the implementation: a plain floating-point scan of all faces (`m9_k4_scan.py`) gives the same maxima to seven digits for every pair, and at each of 797,893 randomly chosen faces the exact critical point lies in its enclosure, with error at most 29% of the bound, and its exact value lies below the computed upper bound (`m9_k4_rig_check.py`). Coordinate ascent over decoders with $k = 5, 6$, with exact polishing of the best points (`j2_cert_polish.py`), finds no certificate either. Nor does a global search over the decoders of the shape of the $m = 8$ certificates, $F(z) = G(z_k, n)$ with n the number

of $i < k$ with $z_i = c$ and all rows of the dangerous sign: Γ_ν is a quadratic in the $2k$ values of G , and enumerating all faces of $[-1, 1]^{2k}$ gives a negative maximum for every $k \leq 8$ (`m9_struct_search.py`, floating point); the same search recovers the certificates of $(-\frac{5}{32}, \frac{1}{16}, \frac{1}{16}, -\frac{1}{32})$ and g_3 . Decoders with small influences do not help either. For the threshold decoders $P = \mathbf{1}[\#\{r : X_r = c\} \leq \tau]$ on k rows of the dangerous sign, the central limit theorem gives $\sigma\Gamma_\nu \rightarrow G_\infty(p) = \sigma h(2p - 1, 2p - 1) + 4|b|(\Pr[G_1 \leq t, G_2 \leq t] - p^2)$ as $k \rightarrow \infty$ and $\mathbb{E}P \rightarrow p$, where G_1, G_2 are standard Gaussians with correlation $-a$ and $\Phi(t) = p$; for all 17 pairs $\min_p G_\infty > 2.8 \cdot 10^{-3}$ (`m9_gauss_limit.py`, floating point). By the invariance principle for correlated spaces [12], the same limit bounds $\mathbb{E}[P(X)P(Z)]$ from below, up to $o(1)$, for every decoder on rows of the dangerous sign whose label coordinates all have influence $o(1)$ (the pair (X_r, Z_r) has maximal correlation a). So, granting these floating-point minima, for each pair there is $\tau > 0$ such that every certificate using only rows of the dangerous sign has a label coordinate of influence at least τ , as the two certificates for $m = 8$ do. The same argument allows a few influential rows. Let A be a set of k_0 label rows, of any signs, and let every other label coordinate have small influence. Condition on the labels $u, v \in \{-1, 1\}^A$ seen in the two columns, let p_u be the conditional mean of P given $X_A = u$ and $\Phi(t_u) = p_u$; conditioning multiplies influences by at most $1/\min_u \Pr[X_A = u]$, so the conditional decoders still have small influences. The lower bound of the invariance principle [12, Theorem 1.14 and Proposition 1.15], which allows different correlated spaces in different coordinates (here all with maximal correlation at most a), gives $\mathbb{E}[P(X)P(Z) \mid X_A = u, Z_A = v] \geq \Pr[G_1 \leq t_u, G_2 \leq t_v] - o(1)$, where $\text{Corr}(G_1, G_2) = -a$. Hence $\sigma\Gamma_\nu \geq V(p) - o(1)$ with

$$V(p) = \sigma h(2\bar{p} - 1, 2\bar{p} - 1) + 4|b| \left(\sum_{u,v} w(u, v) \Pr[G_1 \leq t_u, G_2 \leq t_v] - \bar{p}^2 \right), \quad \bar{p} = \sum_u \Pr[X_A = u] p_u,$$

where w is the joint law of (X_A, Z_A) . Threshold decoders $P = \mathbf{1}[S_X \leq t_{X_A}]$, with S_X a normalised count over many further rows of the dangerous sign, attain $V(p)$ in the limit. Also, $V(p)$ is at most the value of the decoder on the rows A alone with values p_u , since $\Pr[G_1 \leq s, G_2 \leq t] \leq \Phi(s)\Phi(t)$ for negative correlation. We minimised V over $[0, 1]^{2^{k_0}}$ for all 17 pairs, all $k_0 \leq 5$ and all sign patterns of A , using L-BFGS-B with an analytic gradient. The starting points were 150 random points, the minimisers found for $k_0 = 1$ (embedded by adding an ignored row) and the best 0/1 points (all of them for $k_0 \leq 4$). Every minimum found is positive, the smallest being $7.17 \cdot 10^{-5}$ (`m9_mixed_limit.py`, floating point; the minimisation is not certified to be global). For $k_0 \leq 4$ the minima equal the bounds of Proposition 11.16 for every pair except $h = (-\frac{5}{16}, \frac{9}{64}, \frac{9}{64}, -\frac{1}{8})$, where a single row plus a Gaussian part gives $7.84 \cdot 10^{-4}$ instead of $8.11 \cdot 10^{-4}$. Moving to $k_0 = 5$ improves no pair. So, granting these minima, a certificate needs more than five label rows that do not have small influence. By Remark 10.10, the bound of Proposition 10.6 is optimal for the two inequalities used in its proof. Numerically, a linear program over the level weights that also uses the bounds $-\min(p, 1 - p)^2 \leq \mathbb{E}[P(X)P(\tilde{Z})] - p^2 \leq p - p^2$ for product couplings whose correlations in the two kinds of rows range over a 31×13 grid is never weaker than $\Phi^\sharp$ on the grid $p \in 0.005\mathbb{Z}$, is stronger away from the minimiser (by at most 10^{-3}), and has the same minimum over that grid, for all 17 pairs (`coupling_lp_m9.py`). For four pairs, adding the three-copy inequalities $\mathbb{E}[P(X)P(Y)] + p \geq \mathbb{E}[P(X)P(W)] + \mathbb{E}[P(Y)P(W)]$ (from $\mathbb{E}[P(X)P(Y)(1 - P(W))] \geq 0$ and $\mathbb{E}[(1 - P(X))(1 - P(Y))P(W)] \geq 0$; correlation triples on a finite grid) does not change the value at the minimiser of $\Phi^\sharp$ (`coupling_lp3.py`, output `coupling_lp3_m9.txt`). For two of the pairs θ is attained at $\ell = 1$; for them $\Phi^\sharp$ is negative only strictly inside $(q^2, q(1 - q))$, and the bound of Proposition 10.6 is attained by actual decoders at both ends of this interval (Remark 10.10), where $\Phi^\sharp$ is positive. Deciding these candidates needs either a certificate with $k \geq 5$ or an argument that goes beyond the level-weight relaxations tried here; $m = 9$ is the first case with $|J| \leq 2$ where our methods leave candidates open.

Decidability. By (M1), $\mathbb{E}[h(F(C_{\cdot j}))_{j \in J}] = \mathbb{E}[h(B)]$, where $B \in \{-1, 1\}^J$ is obtained by rounding the values $F(C_{\cdot j})$ independently ($B_j = 1$ with probability $\frac{1+F(C_{\cdot j})}{2}$). So criterion (d) of Theorem D asks whether the label source, whose columns $C_{\cdot j}$ are seen by $|J|$ separate parties, lets the parties produce, each applying the same randomised map to its own column, a joint law λ of B with $\sigma \sum_w \lambda(w)h(w) < 0$: a non-interactive simulation problem with identical strategies. Rows of both signs can be combined into one source, a pair of rows of opposite signs per coordinate, since a decoder may ignore rows. For two parties with arbitrary deterministic strategies, De, Mossel and Neeman [3] (after Ghazi, Kamath and Sudan [6] for targets on $\{0, 1\}^2$) showed that every joint law that can be simulated from some number of copies of a finite source is approximated within any $\varepsilon > 0$ by one that can be simulated from $n_0(\varepsilon)$ copies, with n_0 computable. If this extends to identical randomised strategies (and, for $|J| \geq 3$, to $|J|$ parties), then whether $s_g > \text{triv}_g$ is decidable whenever $\inf \sigma \mathbb{E}[h(F(C_{\cdot j}))_j]$ over all k, ν, F is not 0: compute the infimum to a precision below its modulus. The remaining case, infimum exactly 0, includes all 288 boundary functions for $m = 5$, for which the dictator decoder gives the value 0; for the 34 open candidates for $m = 9$ the decoders with $k \leq 4$ stay at least $7 \cdot 10^{-5}$ away from 0. We have not checked these extensions, and the bounds $n_0(\varepsilon)$ are far too large for computation.

Exact value of s_{E2} . Theorem B reduces it to the labelled problem; we only have $0.6251187 < s_{E2} \leq s_{E2}^U < 0.626951$ and $s_{E2}^U > 0.6253166$ (Theorem A(b)–(d); all computer-assisted). Whether $s_{E2} = s_{E2}^U$ is open. The natural label constructions do not close the gap between $\frac{3}{8}J$ and J : for families with two roles, $Q_z = \text{sgn}$ or $Q_z = -\text{sgn}$ according to a function $f(z)$, each γ_ν is a mixture of $\gamma_q(\pm x)$ and $\gamma_{(\text{sgn}, \text{sgn}, \text{sgn})}(\pm x)$ (the latter is at most -0.052), and an exhaustive search over all $f: \{-1, 1\}^k \rightarrow \{0, 1\}$ with $k \leq 4$ finds the largest value $\frac{5}{8} + \frac{3}{8}J$ (up to $6 \cdot 10^{-9}$), attained by $f(z) = z_1$ (Proposition 4.2). The output patterns ν for which the number of columns with role sgn is not determined contribute at most $7.3 \cdot 10^{-5}$ for any f with $k \leq 4$, and every f that gains there loses more elsewhere (`e2_label_roles.py`, floating point). Remark 4.10 locates the losses of Theorem A(d): a sharper pointwise bound would have to use the joint structure of the three test functions, and a majorant that is not a sum of functions of single coordinates could avoid paying for points where $\Xi < 0$ (the ridge majorants tried in Remark 4.10 give no improvement).

One-dimensionality (second part of Q5). For E2 the value s^L is attained by every q in every dimension (Theorem 3.2); the constructions above for s^U and s^{lab} are one-dimensional, but we do not know whether the suprema are attained in dimension one. Some numerical evidence: for halfspaces $q_j = \text{sgn}\langle v_j, \cdot \rangle$ with unit normals $v_j \in \mathbb{R}^n$, $\text{val}(q_1, q_2, q_3)$ depends only on the Gram matrix of the v_j , and given $\mathcal{G}_0 = x$ the variables $U_j = \langle v_j, \mathcal{G}_j \rangle$ are normal with means $\rho\langle v_j, x \rangle$ and covariance matrix $I - \rho^2(\langle v_i, v_j \rangle)_{ij}$. Every perturbation of the collinear configuration $(\text{sgn}, \text{sgn}, -\text{sgn})$ lowers the moduli $|\langle v_i, v_j \rangle|$ by some $\delta_{ij} \geq 0$. Computing val with exact bivariate orthant probabilities (Owen’s T function) and adaptive quadrature, the change is $D_{12}\delta_{12} + D_{13}\delta_{13} + D_{23}\delta_{23}$ to first order, with $D_{12} \approx -4.4 \cdot 10^{-4}$ and $D_{13} = D_{23} \approx -2.1 \cdot 10^{-4}$ (finite differences at $\varepsilon = 0.1, 0.05$ with Richardson extrapolation); for instance, rotating v_3 by 0.1 lowers val from 0.6253166 by $2.1 \cdot 10^{-6}$ (`e2_halfspace_angles.py`, output `e2_halfspace_angles.txt`). So among halfspaces the one-dimensional configuration is a strict local maximum (numerically). Among one-dimensional mean-zero step functions q_j (bins of width 0.1 on $[-6, 6]$), projected gradient ascent on a smoothed objective started at $(\text{sgn}, \text{sgn}, -\text{sgn})$ finds no improvement, perturbed starts reach at most $\mathbb{E}\gamma^+ = 2.43 \cdot 10^{-4}$ and random starts stay at 0 (a fixed number of steps, without a convergence test; `e2_sU_opt1d.py`, floating point).

A Verification scripts

All scripts are in the directory `numerics/` (Python 3 with `numpy`, `scipy`, `sympy` and `mpmath`; `e2_lower_rig.py`, `e2_upper_rig.py`, `orbit_k3_exact.py`, `m9_k4_rig.py` and `m9_k4_rig_check.py` also need `python-flint`); their outputs are in `numerics/outputs/`.

- `e2_rigorous.py`: exact certificate for Theorem 3.2 (Sturm count on the rational cubic $\frac{1}{2} - \bar{B}$) and the integral I .
- `e2_lower_rig.py`: the computer-assisted part of Proposition 4.3 (formula (2) checked against Gauss–Hermite quadrature; rigorous enclosure of $\int_{x_1}^{12} \gamma_q \varphi dx$ with Arb; output `e2_lower_rig.txt`).
- `e2_upper_rig.py`: the computer-assisted part of Theorem 4.5 (Remark 4.9): the constant K_3 (exact), construction of ψ by a linear programme, verification of the majorant property on all cells with interval arithmetic, convexity and $\mathbb{E}\psi(U)$ in Arb ball arithmetic (`python3 e2_upper_rig.py 400`; outputs `e2_upper_rig_400.txt`, `e2_upper_psi_400.txt`, `e2_upper_z_400.txt`).
- `e2_upper_sanity.py`: floating-point tests of Lemma 4.6 on halfspaces and unions of intervals, and the losses along $(\text{sgn}, \text{sgn}, -\text{sgn})$ quoted in Remark 4.10; `e2_upper_ridge.py`: the linear programme with ridge majorants of Remark 4.10 (floating point). `e2_upper_maj.py`, `e2_upper_herm.py`, `e2_upper_cells.py`: floating-point precursors (other pointwise bounds; the closed form of Lemma 4.7 compared with a brute-force maximisation over the ellipsope; the cell linear programme).
- `q6_exact.py`: exact classification for $m \leq 4$ (Theorem C(v)).
- `q6_d4r1.py`: exact enumeration for $m = 5$, $|J| = 4$ (Proposition 11.1).
- `q6_hsearch.py`: enumeration of candidate h for given $(|J|, |U|)$ with TRIV/L0/LAB1 tests (floating point on dyadic data, hence exact; see the proof of Proposition 11.1; outputs `q6_hsearch_d2_r3.txt`, `q6_hsearch_d3_r2.txt`).
- `q6_realise_check.py`, `q6_exact_k2.py`, `q6_hard288.py`: exact checks on the 348 candidates (realisability, $\|h\|^2 = |\mu|$, $k \leq 2$ certificates) and the list of the 288 boundary functions.
- `g0_check.py`: every number in the proof of Proposition 9.1 (exact). `q6_g0.py`, `q6_g0_variation.py`: data of g_0 (exact) and Boolean decoders with $k \leq 2$; first variation at the dictator decoder.
- `hard288_orbits.py`: orbits of the 288 boundary functions under the symmetries of Lemma 11.2, with a check that the list is closed under them.
- `khintchine_q_check.py`: exact checks of the constants of the proof of Theorem K for $q = \frac{3}{8}, \frac{2}{5}, \frac{5}{12}$ (the table of §9.4, the identities for general q , the case computations and the sign pattern of p_q); a floating-point scan over q (grid of step 0.005); $\mathbb{E}|L|$ for random unit vectors and a local search away from dictators.
- `khintchine_interval.py`: the proof of Theorem K for all $q \in [\frac{3}{8}, \frac{1}{2}]$ (exact rational arithmetic): the thresholds a_A satisfying (11) on 64 intervals of q , and the bisection proving (13) on the Case C ranges (249 boxes); it also checks that $P(q, \cdot)$ is a positive multiple of the tabulated p_q .

- `e2_label_roles.py`: exhaustive search over two-role label families for E2 ($Q_z = \pm \text{sgn}$), $k \leq 4$ (floating point, Owen's T and grid quadrature; §11).
- `khintchine_threshold.py`: the counterexample $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ for $q < 1 - \frac{1}{\sqrt{2}}$ (enumeration) and a floating-point search for the largest $\mathbb{E}|L|/(2\tau)$ over unit vectors with $n \leq 7$ on a grid of $q \in [0.28, 0.375]$ (Remark 9.8(c)).
- `g1_check.py`: exact data of g_1 and every number in the proof of Proposition 9.10; brute-force comparison of $\Gamma_\nu(F)$, the reduction bound and the spectral formula for decoders with $k \leq 4$.
- `d12_check.py`: the eigenvalue identities behind Theorem L for general q (sympy), the identity (15) on random $f \geq 0$ and its equality cases; exact data of g_2 and every number in the proof of Proposition 9.11; brute-force comparison of $\Gamma_\nu(F)$, the reduction bound, the spectral formula and $\frac{1}{2}\mathbb{E}[P^2 - P]$ for decoders with $k \leq 4$ (exact for $k \leq 2$); the claims of §11 about the one-class orbits B12 and C12.
- `level1_profile_test.py`: floating-point test of $W^1(f) \leq qp + \frac{1-2q}{q}p^2$ through the tail functional of unit linear forms ($q = \frac{2}{5}$, $n \leq 5$), run before Theorem L was found.
- `thmL_sharpness.py`: the equality cases of Theorem L, exactly, for all $\{0, 1\}$ -valued f with $n \leq 4$ and for random nonnegative f of degree at most 2 with $n = 4$ ($q \in \{\frac{1}{5}, \frac{3}{10}, \frac{1}{3}, \frac{2}{5}\}$), including $(\mathbb{E}f)^2/\mathbb{E}f^2 \in [q^2, q]$; the formulas for ANDs of t coordinates (exact); (16) on test functions (floating point).
- `oneclass_analysis.py`, `orbit_coord_ascent.py`: exact data ($\pi_\pm$, $\beta_\pm$, correlations, $k = 1$ values) of the five one-class orbits; coordinate ascent (exact one-dimensional steps, floating point) over single decoders $F: \{-1, 1\}^k \rightarrow [-1, 1]$ for representatives of all orbits except D12.
- `khintchine_analytic_check.py`: floating-point sanity checks of the analytic proof of Theorem K for $q = \frac{3}{8}$ (majorant (12), the functions of Cases A–C, the cubic $p_q(s)$, and $\mathbb{E}|L|$ for random unit vectors by exact enumeration).
- `khintchine_certificates.py`: independent computer-assisted proof of Case C of Theorem K (exact rational dual certificates on 560 cells; square roots enclosed by rational bounds).
- `khintchine_L1_test.py`, `khintchine_moment_lp.py`: numerical exploration of Theorem K (gradient ascent over unit vectors with exact enumeration, $n \leq 10$; the moment-problem relaxation of the inductive step).
- `g0_pairwise_test.py`, `g0_coord_ascent.py`, `level1_bellman.py`: numerical tests for g_0 (the pairwise inequality of Lemma 9.2; coordinate ascent for labelled certificates with $k \leq 6$; a Bellman-function approach to the level-one bound, which is too weak).
- `e2_check.py`, `polymorph.py`, `e2_sU_opt1d.py`: numerical values for E2 (tensor Gauss–Hermite quadrature for $\gamma(x)$, adaptive quadrature in x ; one-dimensional optimisation for s_{E2}^U over step functions, with the x -integral on a 361-point grid; the discretised value of $(\text{sgn}, \text{sgn}, -\text{sgn})$ is 0.6253169, against the exact 0.6253166, and the optimiser does not improve on it).
- `e2_halfspace_angles.py`: val for E2 and triples of halfspaces with arbitrary normals (exact bivariate orthant probabilities via Owen's T , Gauss–Legendre in the first coordinate, adaptive quadrature in x ; floating point).

- `q6_deeplabels.py`, `q6_contrlabeleds.py`, `g0_finite_x.py`: numerical searches (floating point) for labelled certificates at $x = \infty$ and at finite x .
- `gF_check.py`: exact data of g_F (Fourier coefficients, dependence on all coordinates, h , $\pi_{\pm}$, the correlations, ρ_{12} and Φ) for Proposition 10.3.
- `maxcorr_J2.py`: the exact classification of Proposition 10.7, including the tests of Propositions 10.4(i) and 10.6 and the exact search for certificates with $k \leq 3$ (`python3 maxcorr_J2.py r [workers]`, $r = 3, 4, 5, 6$, and $r = 7$ for §11; outputs in `numerics/outputs/maxcorr_J2_r*.txt`).
- `m9_open.py`: the exact data of the 34 candidates left open for $m = 9$ (§11): θ , the negative minimum of $\Phi^{\sharp}$ and its location, the maximum of Γ_{ν} for $k \leq 3$, and the AND decoder for $\ell = 1$.
- `j2_cert_polish.py`: coordinate ascent over decoders with exact polishing (floating point search; the 17 pairs for $m = 9$ with $k = 4, 5, 6$). `coupling_lp3.py`: the linear programme with two- and three-copy constraints (floating point); `coupling_lp_m9.py`: its two-copy version for all 17 pairs on a grid of p , compared with $\Phi^{\sharp}$. `formula_envelope.py`, `split_bound.py`: other explorations for $m = 9$ (read-once formula decoders; recursive splitting bounds), which decided nothing.
- `fplus_check.py`: the worked examples of Proposition 10.4 (exact).
- `fsharp_check.py`: the minima of $\Phi^{\sharp}$ for the 18 candidates in $F^{\sharp}$ and the worked example of Proposition 10.6 (exact).
- `j2_k2_exact.py`: the exact maximum of Γ_{ν} over $[-1, 1]$ -valued decoders with given k and ν (face enumeration, as in the proof of Proposition 10.9).
- `m8_check.py`: the function g_3 and the second nontrivial pair for $m = 8$; all numbers in Proposition 10.9 and in the paragraph before it (exact).
- `fsharp_sanity.py`, `sharp_test.py`, `tcoupling_test.py`, `fpp_test.py`, `coupling_lp_J2.py`, `coupling_lp_dual.py`: floating-point tests of Lemma 10.5 and Proposition 10.6 on random decoders, of other choices of the couplings, and of the linear programme over finite grids of couplings (which gave no improvement on these candidates).
- `j2_cert_search.py`: ascent over single decoders with exact re-evaluation, which found the certificates for $m = 8$.
- `j2_ascent.py`: coordinate ascent over single decoders for $|J| = 2$ (g_F , the $m = 6$ candidate not decided by Theorem F, and the 11 pairs for $m = 8$ left by Proposition 10.4).
- `class_flips.py`: level-one and cubic coefficients of the 14 orbit representatives and the class flips of Lemma 11.3 (Table 1).
- `thmF_boundary.py`: exact certificates that the hypothesis of Theorem F fails for `rep3` and `rep8`.
- `classwise_ascent.py`: coordinate ascent over class-wise decoders for the nine two-class orbits.

- `rep3_star.py`, `cert_sdp.py`: alternating maximisation for (30), $k \leq 7$; the script reports the maximum of $2\mathbb{E}[H(C_2)G(C_3)] - \mathbb{E}[G(C_1)G(C_3)]$ over $G = 2P - 1$ and $H: \{-1, 1\}^k \rightarrow [-1, 1]$, which is 1 plus four times the maximum of the left side minus the right side of (30) (output `rep3_star.txt`); the SDP certificate search for rep3 described after (30) (CLARA-BEL through `cvxpy`; output `cert_sdp_rep3.txt`), and the same search for rep5 (output `cert_sdp_rep5.txt`).
- `rep3_coupling.py`: exact check of every number in the proof of Proposition 11.5 (h' , $\tilde{\pi}_\pm$ and their pair marginals, ϕ) and exact random tests of $\Gamma_\nu(F) \leq \mathbb{E}[h'(Y)] = -\frac{1}{8}\mathbb{E}[\phi] \leq 0$ for $k \leq 3$ (output `rep3_coupling.txt`). `coupling_search.py`: exact decision, for the nine two-class orbits, of whether the re-coupling argument of Proposition 11.5 applies (Fourier–Motzkin elimination; output `coupling_search.txt`).
- `rep2_coupling.py`: exact check of every number in the proof of Proposition 11.6 (h' , $\rho_\pm$ and their pair marginals, ψ , the flip to the orbit of h_{g_0}) and exact random tests of $\Gamma_\nu(F) \leq \mathbb{E}[h'(Y)] = -\frac{1}{8}\mathbb{E}[\psi] \leq 0$ for $k \leq 3$ (output `rep2_coupling.txt`).
- `rep1_coupling.py`, `rep7_coupling.py`: the same for Propositions 11.8 and 11.10 (h' and R , $\rho_\pm$ and their pair and triple marginals, $R \leq 0$, the class flips to the orbits of rep9 and of h_{g_1} , and exact random tests for $k \leq 3$; outputs `rep1_coupling.txt`, `rep7_coupling.txt`).
- `rep6_coupling.py`: the same for Proposition 11.12 (h' , $\rho_\pm$ and their pair and triple marginals, the matching of the coefficients of R and h' , $R \leq 0$, the flip to the orbit of C12, and exact random tests for $k \leq 3$; output `rep6_coupling.txt`).
- `coupling_designs.py`: the search over coupling designs described before Conjecture 11.14 (floating-point linear programming, designs up to permutations of the columns within a class; outputs `coupling_designs_pairs.txt`, `coupling_designs_cubic.txt` for $N \leq 5$ and `coupling_designs_cubic6.txt` for $N = 6$).
- `milp_designs.py`: the mixed-integer formulation of the design search for a fixed number of columns and a fixed assignment of the columns to the classes, for one-class and two-class h , with the optional spectral inequalities for one class (floating point, HiGHS through `scipy.optimize.milp`; outputs `milp_designs_run1.txt`, `milp_designs_run2.txt`, `milp_designs_small.txt`). `oneclass_designs.py`: the enumeration of one-class designs (floating-point linear programming). `eps_designs.py`: the least ε with $R \leq \varepsilon$ over designs with N columns (same formulation; outputs `eps_designs_B12.txt`, `eps_designs_D12.txt`, `eps_designs_2class.txt`, and the interrupted search `eps_designs_rep10_N8_partial.txt`, whose runs all reached the time limit).
- `rep5_reduction.py`: exact checks for the paragraph “A reduction for rep5” (h' and $4h'(y) = -1 - y_1(1 + y_2 + y_3)$, the two pair laws and their coincidence with those of Proposition 9.11, $\Gamma_\nu = 0$ at the two decoder pairs, the equality cases of (28) and $\Gamma_\nu = -\frac{1}{100}$ at the third, the enumeration of the $\{0, 1\}$ -valued P with $k \leq 3$, and exact random tests for $k \leq 3$ of $\Gamma_\nu \leq \mathbb{E}[h'(Y)]$, of the formula for $\mathbb{E}[h'(Y)]$, of the best response in F_s , of $\Psi(P, P)$ against the expression of Proposition 9.11 and of the minimum of Ψ over B ; output `rep5_reduction.txt`). `rep5_relax.py`: the alternating maximisation for (28) and its analogues for the other cubic-free majorants (floating point; output `rep5_relax.txt`). `rep5_scalar.py`: the level-weight relaxation of Ψ and its minimum on a grid (floating point; output `rep5_scalar.txt`).

- **rep510_exact.py**: exact checks for the paragraph “Exact forms for rep5 and rep10” (the Fourier expansions of the two representatives, the identities for $4\max_{F_s} \Gamma_\nu$ against a brute-force maximum over all $\{-1,1\}$ -valued F_s for 56 random P per orbit with $k \leq 3$ and every number of positive labels, and the equality cases; output **rep510_exact.txt**). **rep510_ascent.py**: the alternating maximisation of Γ_ν for rep5 and rep10 with $k \leq 8$ (floating point; outputs **rep510_ascent_rep5.txt**, **rep510_ascent_rep10.txt**, the positive control **rep510_ascent_control.txt**, and **rep510_ascent_many_k5.txt** with 200 random starts per sign pattern for $k \leq 5$). **rep510_k4_boolean.py**: both one-function inequalities for all 2^{16} $\{0,1\}$ -valued P with $k = 4$ and every ν , in exact integer arithmetic (row laws scaled to integers), with the list of equality cases and random cross-checks against **rep510_exact.py**; output **rep510_k4_boolean.txt**. **rep10_majorant.py**: exact checks for the paragraph “A reduction for rep10” (the zeros of π_+ , the bound $t \leq 1$, the expansion of h'' and its vertex form, the third form for 56 random $[0,1]$ -valued pairs (P,B) with $k \leq 3$, the identity for $4\max_{F_s} \mathbb{E}[h''(Y)]$ against a brute-force maximum over F_s for 56 random P with $k \leq 3$, and (29) for all $\{0,1\}$ -valued P with $k = 4$ and every ν in integer arithmetic, with the equality cases; output **rep10_majorant.txt**). The alternating maximisation for h'' is **rep510_ascent.py** with the objective passed in the variable OBJ (floating point; outputs **rep10_majorant_ascent.txt** for $k \leq 8$ and **rep10_majorant_ascent_many_k5.txt**). **majorant_k3_exact.py**: the face enumeration of Proposition 11.4 with the integrand h'' (rep10), $h + \mathbf{1}_{+-}$ or $h + \mathbf{1}_{+++}$ (rep5) and the row laws of h , for $k \leq 3$ (python-flint; for $k \leq 2$ every maximum over F_p is checked against the Fraction-based enumeration **cube_max** of **j2_k2_exact.py**, and for every vertex F_s the quadratic form in F_p is checked against a direct evaluation of $\mathbb{E}[h'(Y)]$ at a random rational F_p), and (28) for all $\{0,1\}$ -valued P with $k = 4$ in integer arithmetic, with the equality cases and random cross-checks against a Fraction computation (output **majorant_k3_exact.txt**). **rep510_k4_half.py**: the four one-function inequalities for all 3^{16} $\{0, \frac{1}{2}, 1\}$ -valued P with $k = 4$ and every $\nu = (+1)^r(-1)^{4-r}$ (which suffices by permuting the labels), in integer arithmetic (values of $2P$), with the number of equality cases and of those that take the value $\frac{1}{2}$ (none), and 10 random cross-checks per orbit and ν against Fraction computations (those of **rep510_exact.py** and **rep10_majorant.py**, and a direct one for (28)) (output **rep510_k4_half.txt**). **rep510_boolean.py** with the C kernel **rep510_boolean.c**: the four one-function inequalities for all 2^{32} $\{0,1\}$ -valued P with $k = 5$, and again all 2^{16} with $k = 4$, for every $\nu = (+1)^r(-1)^{k-r}$ (which suffices by permuting the labels). It enumerates P in Gray-code order, in 8 ranges per ν , updating all accumulators exactly in integer arithmetic (every value has modulus at most $9D$, and $9D < 2^{31}$ is asserted), and at the end of each range recomputes them from scratch and compares. The $k = 4$ results are compared with a direct evaluation, and sample points of every range (at pseudo-random Gray indices) with Fraction computations from **rep510_exact.py**. The script asserts that the equality cases are exactly the families listed in the paragraph “A reduction for rep10” (output **rep510_boolean.txt**).
- **oneclass_certs.py**: exact check of the one-class certificates for A24 and C12 (paragraph “One-class designs”: h' , the laws and all their assigned marginals, the matching of the coefficients of R and h' by law, $R \leq 0$, and exact random tests of the chains for $k \leq 3$; output **oneclass_certs.txt**).
- **b12_cert.py**: exact check of the seven-column certificate of Proposition 11.15 (properties (a)–(c) of $\rho_\pm$, the matching of the coefficients of R and h by law, the formula for R , the elimination of a_4, a_5 , $R \leq 0$ at the 128 vertices, and exact random tests of $\Gamma_\nu(F) = \mathbb{E}[R] \leq 0$

for $k \leq 3$; output `b12.cert.txt`).

- `m9_struct_search.py`: global search (face enumeration, floating point) over decoders $F(z) = G(z_k, n)$ for the 34 open $m = 9$ candidates, $k \leq 8$, validated on the two $m = 8$ certificates (output `m9_struct_search.txt`).
- `m9_k4_rig.py`: the computation of Proposition 11.16 (`python3 m9_k4_rig.py 4 8`, four minutes with eight processes; output `m9_k4_rig.txt`); run with $k = 3$ it reproduces the exact maxima of `m9_open.py`, and on the two nontrivial $m = 8$ pairs it reports the exact values of the certificates of Proposition 10.9 ($\frac{4317675409}{14846652293120}$, run with $k = 3$) and of the paragraph before it ($\frac{67561}{411505920}$, run with $k = 2$). `m9_k4_rig_check.py`: comparison of its enclosures with exact rational solutions at random faces (output `m9_k4_rig_check.txt`). `m9_k4_scan.py`: the same face enumeration in plain floating point (output `m9_k4_scan.txt`).
- `m9_gauss_limit.py`: the Gaussian limit G_∞ of threshold decoders for the 17 open $m = 9$ pairs (floating point; output `m9_gauss_limit.txt`).
- `m9_mixed_limit.py`: minimisation of the mixed limit V ($k_0 \leq 5$ rows read exactly plus a Gaussian part) for the 17 open $m = 9$ pairs (floating point, multistart; output `m9_mixed_limit.txt`).
- `orbit_kkt.py`: the first-order conditions at the dictator decoder for the 14 orbit representatives, and on the flat face of `rep10` the constants $A(s)$, $\alpha_j(s)$ and the brackets $\frac{1}{12}$, $\frac{1}{16}$ that make them strict for every k (exact rational arithmetic; output `orbit_kkt.txt`).
- `orbit_k2_exact.py`, `orbit_k3_exact.py`: the computation of Proposition 11.4, the exact maximum of Γ_ν over all $[-1, 1]$ -valued class-wise decoders for the nine two-class orbit representatives, $k \leq 2$ (Python fractions) and $k = 3$ (python-flint); the second script is checked against the first, vertex by vertex, for $k \leq 2$ (outputs `orbit_k2_exact.txt`, `orbit_k3_exact.txt`).

None of the scripts is needed for the proofs of Theorems B, D, E, F and L, of Theorem K for $q \in \{\frac{3}{8}, \frac{2}{5}, \frac{5}{12}\}$ (every number is displayed in §9.4), and of Propositions 9.10, 9.11, 10.3, 11.5, 11.6, 11.8, 11.10, 11.12 and 11.15, which are analytic; the scripts listed for them are cross-checks. The computer-assisted results are: Theorem A(a) (exact Sturm counts, `e2_rigorous.py`); Theorem A(b),(c) (Proposition 4.3, Arb enclosure, `e2_lower_rig.py`); Theorem A(d) (`e2_upper_rig.py`); Theorem C(v) (`q6_exact.py`, exact); Theorem K for general $q \in [\frac{3}{8}, \frac{5}{8}]$ (`khintchine_interval.py`, exact rational arithmetic); Propositions 10.7 and 10.9 and Corollary 10.8 (`maxcorr_J2.py`, `j2_k2_exact.py`, `m8_check.py`, exact); Proposition 11.1 (`q6_hsearch.py`, `q6_exact_k2.py`, `q6_realise_check.py`, exact; `q6_d4r1.py` is a cross-check); Proposition 11.4 (`orbit_k2_exact.py`, `orbit_k3_exact.py`, exact); and Proposition 11.16 (`m9_k4_rig.py`, floating point with rigorous error bounds and an exact fallback).

References

- [1] C. Borell. Geometric bounds on the Ornstein–Uhlenbeck velocity process. *Z. Wahrsch. Verw. Gebiete* 70(1):1–13, 1985.
- [2] G. Chase, Y. Filmus, D. Minzer, E. Mossel, N. Saurabh. Approximate polymorphisms. *STOC 2022*; arXiv:2106.00093. (Numbering of statements as in the arXiv version.)

- [3] A. De, E. Mossel, J. Neeman. Non interactive simulation of correlated distributions is decidable. SODA 2018, 2728–2746.
- [4] R. Eldan. A two-sided estimate for the Gaussian noise stability deficit. *Invent. Math.* 201(2):561–624, 2015.
- [5] E. Friedgut. On the measure of intersecting families, uniqueness and stability. *Combinatorica* 28(5):503–528, 2008.
- [6] B. Ghazi, P. Kamath, M. Sudan. Decidability of non-interactive simulation of joint distributions. FOCS 2016, 545–554.
- [7] L. Goldstein, G. Reinert. Stein’s method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.* 7(4):935–952, 1997.
- [8] C. Jones. A noisy-influence regularity lemma for Boolean functions. *arXiv:1610.06950*, 2016.
- [9] P. Ivanišvili, D. Li, R. van Handel, A. Volberg. Improving constant in end-point Poincaré inequality on Hamming cube. *arXiv:1811.05584*, 2018.
- [10] P. Ivanišvili, Y. Stone. Sharpening the gap between L^1 and L^2 norms. *arXiv:2407.04835*, 2024.
- [11] L. Yu. A sharp small-coefficient variant of Khintchine’s inequality and the sharp $\pi/2$ theorem. *arXiv:2608.29703*, 2026.
- [12] E. Mossel. Gaussian bounds for noise correlation of functions. *GAFA* 19(6):1713–1756, 2010.
- [13] E. Mossel, J. Neeman. Robust optimality of Gaussian noise stability. *J. Eur. Math. Soc.* 17(2):433–482, 2015.
- [14] R. O’Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.
- [15] K. Oleszkiewicz. On a nonsymmetric version of the Khinchine–Kahane inequality. In: *Stochastic Inequalities and Applications*, *Progr. Probab.* 56, Birkhäuser, 2003; doi:10.1007/978-3-0348-8069-5_11.
- [16] Y. Stone. The Poincaré inequality on the Hamming cube for vector-valued functions. PhD thesis, University of California, Irvine, 2025 (eScholarship, escholarship.org/uc/item/2wv0w7cz).
- [17] R. van Handel. Private communication, quoted in [10].
- [18] H. S. Witsenhausen. On sequences of pairs of dependent random variables. *SIAM J. Appl. Math.* 28(1):100–113, 1975.