AND testing with square-root loss: the optimal ε–δ dependence is $\Theta(\sqrt{\varepsilon})$, and polynomial bounds for all approximate polymorphisms

Paata Ivanisvili

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

21 October 2026

Abstract

Let $f\colon \{0,1\}^n \to \{0,1\}$ and let $\varepsilon = \Pr_{\mathbf{x},\mathbf{y}}[f(\mathbf{x}\wedge\mathbf{y}) \neq f(\mathbf{x})\wedge f(\mathbf{y})]$ for independent uniform $\mathbf{x},\mathbf{y}$. Filmus, Lifshitz, Minzer and Mossel (STOC 2020) proved that f is δ-close to a constant or to an AND of a subset of the coordinates, where $\delta = \delta(\varepsilon) \to 0$ independently of n. They asked for the optimal dependence of δ on ε; the best known bound was $\delta = 1/\log^{\Omega(1)}(1/\varepsilon)$ (FLMM, 2021). We prove that

$$\min_{g\in\{0\}\cup\{\mathrm{AND}_S\}} \Pr[f\neq g] \;\le\; \min\Big\{\mathbb{E}f,\; \frac{c_0\,\varepsilon}{\mathbb{E}f}\Big\} \;\le\; \sqrt{c_0\,\varepsilon}, \qquad c_0 = 4+2\sqrt{2} < 6.83,$$

and that $\sqrt{\varepsilon}$ cannot be improved: there are functions with $\varepsilon(f)\le\varepsilon$ that are $(1-o(1))\sqrt{\varepsilon}$-far from all constants and ANDs. Thus the optimal dependence is $\delta^*(\varepsilon) = \Theta(\sqrt{\varepsilon})$, which proves the polynomial conjecture of FLMM with the optimal exponent. The order of the lower bound is elementary: $x_1\cdots x_k(1-x_{k+1})$ is at distance exactly $\sqrt{\varepsilon/2}$, so the bound $\delta = O(\varepsilon)$ conjectured in their 2021 note already fails for these functions. The substance is the matching upper bound. The proof is elementary (no junta theorem, no hypercontractivity). Its main ingredient is a *spectral containment inequality*: if $\mathbf{S}$ is drawn from the Fourier spectral distribution of f and $\mathbf{x}$ is uniform on $f^{-1}(1)$, then $\Pr[\mathbf{S}\subseteq \mathrm{supp}(\mathbf{x})] \ge 1-\varepsilon/(\mathbb{E}f)^2$. The same method gives the sharp bound $\Theta(\varepsilon^{1/m})$ for the m-ary test $f(\mathbf{x}_1\wedge\cdots\wedge\mathbf{x}_m) = \bigwedge_j f(\mathbf{x}_j)$, the bound $\sqrt{6\varepsilon/(1-p)}$ for p-biased inputs, and polynomial ($O_{p,\rho}(\sqrt{\varepsilon})$) bounds for the three-function, two-measure version (Theorem 2.10 of FLMM). The leading constant is not 1: there are functions with $\mathrm{dist}(f)^2 > 1.351\,\varepsilon(f)$ at arbitrarily small ε, so $1.162 < \limsup_{\varepsilon\to 0}\delta^*(\varepsilon)/\sqrt{\varepsilon} \le \sqrt{c_0} < 2.614$. In the relative form the constant is sharp near the ANDs: $\mathrm{dist}(f) \le (\frac{1}{2} + O((\varepsilon/\mu^2)^{1/2}))\,\varepsilon/\mu$ with $\mu = \mathbb{E}f$, and $\frac{1}{2}$ is attained. Finally, the coupling works for approximate polymorphisms of an arbitrary $g\colon \{0,1\}^m \to \{0,1\}$. We prove that every δ-approximate (multi-)polymorphism of g is $O_g(\delta^{c_g})$-close to an exact one, with $c_g > 0$ depending only on m; the previous bounds (Chase–Filmus–Minzer–Mossel–Saurabh, STOC 2022) were of tower type. This answers their Open Question 2: the dependence is always polynomial, and its exponent depends on g. It is linear for every balanced g that depends on at least two coordinates. For unbalanced g let $e(g)$ be the number of coordinates of a constant input that must be flipped to change the value of g, maximized over the two constant inputs. Then the optimal distance of a δ-approximate polymorphism from the skew polymorphisms is $\Theta_g(\delta^{1/e(g)})$ for every non-constant unbalanced g (for instance $e = m$ for AND_m, OR_m, NAND_m, NOR_m); for functions close to constants this uses a strong data processing inequality. So the optimal exponent for polymorphisms is known for every g; for g depending on at least two coordinates, the loss is linear if and only if g is balanced or $e(g) = 1$. For multi-polymorphisms (the functions $f_0,\dots,f_m$ may differ) we also find the optimal exponent: the optimal distance is $\Theta_g(\delta^{1/E(g)})$, where $E(g)$ is the largest k such that fixing $m-k$ of the variables of g can leave an AND of k literals or its negation. So for g depending on at least two coordinates the loss is linear if and only if g is affine;

for instance, for Maj_3 it is linear for polymorphisms and quadratic for multi-polymorphisms. Finally, using the reverse hypercontractivity of Mossel, Oleszkiewicz and Sen, the polynomial bound extends to correlated inputs: for every g and every full-support distribution μ of the rows (the setting of Alekseev and Filmus), the dependence is polynomial, with exponent depending only on m and $\min \mu$. This proves the polynomial part of the Alekseev–Filmus conjecture for every functional predicate. The method also extends to larger alphabets: for $g\colon \Sigma^m \to \Sigma$ such that changes of single coordinates within the connected components of their collision graphs ($a \sim b$ if $g(u^{j\to a}) = g(u^{j\to b})$ for some context u) connect all output values (for instance min, max, every finite semilattice operation, every near-unanimity operation, multiplication modulo q, and the rounded mean $\lfloor (x_1+x_2)/2 \rfloor$; for $q=3$, 87% of the binary operations that depend on both coordinates), the dependence is again polynomial; these functional predicates are not covered by the larger-alphabet theorem of Alekseev and Filmus. For the non-functional predicate $\{(a,b,c) \in \mathbb{Z}_3^3 : a+b+c \neq 0\}$, which they name as the simplest predicate their techniques cannot handle, we prove stability under the uniform distribution on the predicate, with the sharp dependence: distance $(1+o(1))\sqrt{\delta}$. Polynomial stability extends to every full-support distribution on the predicate, with an exponent that must depend on the distribution, which answers their question for this predicate; for the distributions with uniform marginals and smallest atom α, the best exponent valid for all of them lies between $9\alpha(1-9\alpha)$ and 9α, and it equals 9α for $\alpha \le \frac{1}{36}$; near the affine solutions the dependence is linear for every full-support distribution. The optimal exponent $\frac{1}{2}$ and the sharp constant 1 also hold for $\{a \in \mathbb{Z}_3^m : a_1 + \cdots + a_m \neq 0\}$, $m \ge 4$, under the uniform distribution, and polynomial stability holds for these predicates under every full-support distribution, with the best exponent for uniform marginals between $9\alpha'(1-9\alpha')$ and $9\alpha'$, and equal to $9\alpha'$ for $\alpha' \le \frac{1}{36}$, where $\alpha' = 3^{m-3}\alpha$; for $m \ge 3$ these predicates are, up to relabelling the alphabet in each coordinate, exactly the largest predicates over $\mathbb{Z}_3$ without a flexible coordinate.

1 Introduction

For $x, y \in \{0,1\}^n$ write $x \wedge y = (x_1 \wedge y_1, \dots, x_n \wedge y_n)$. For $S \subseteq [n]$ let $\mathrm{AND}_S(x) = \prod_{i \in S} x_i$ (so $\mathrm{AND}_\emptyset \equiv 1$), and let

$$\mathcal{A}_n = \{0\} \cup \{\mathrm{AND}_S : \ S \subseteq [n]\}.$$

The functions $f\colon \{0,1\}^n \to \{0,1\}$ satisfying $f(x \wedge y) = f(x) \wedge f(y)$ for all x, y are exactly the members of $\mathcal{A}_n$ (in social choice: the oligarchies and the constants). Indeed, if $f \not\equiv 0$ is such a function then $f^{-1}(1)$ is closed under $\wedge$ and is an up-set (if $f(x)=1$ and $y \ge x$ then $1 = f(x \wedge y) = f(x)f(y)$), so $f^{-1}(1) = \{x \ge m\}$ where m is the meet of all points of $f^{-1}(1)$, i.e. $f = \mathrm{AND}_{\mathrm{supp}(m)}$. For $f\colon \{0,1\}^n \to \{0,1\}$ define, with $\mathbf{x}, \mathbf{y}$ independent and uniform on $\{0,1\}^n$,

$$\varepsilon(f) = \Pr_{\mathbf{x},\mathbf{y}}\big[f(\mathbf{x} \wedge \mathbf{y}) \neq f(\mathbf{x})f(\mathbf{y})\big], \qquad \mathrm{dist}(f) = \min_{g \in \mathcal{A}_n} \Pr_{\mathbf{x}}[f(\mathbf{x}) \neq g(\mathbf{x})],$$

and

$$\delta^*(\varepsilon) = \sup_{n \ge 1} \ \sup\big\{\mathrm{dist}(f) : \ f\colon \{0,1\}^n \to \{0,1\},\ \varepsilon(f) \le \varepsilon\big\}.$$

Nehama [24] proved $\mathrm{dist}(f) = O((n\varepsilon)^{1/3})$. Filmus, Lifshitz, Minzer and Mossel [11] removed the dependence on n: $\delta^*(\varepsilon) \to 0$ as $\varepsilon \to 0$ (their Theorem 1.1), with an exponential loss, and they asked ([11, Open Question 1]): *"What is the optimal dependence between ϵ and δ in Theorem 1.1?"*, conjecturing a polynomial dependence. In a 2021 note [12] they gave a simpler proof with $\delta^*(\varepsilon) \le 1/\log^{\Omega(1)}(1/\varepsilon)$, based on a Fourier tail bound and the junta theorem of Kindler–Kirshner–O'Donnell [18], and conjectured that their theorem holds with $\delta = O(\varepsilon)$. Read literally, this fails for an elementary reason. The function $f = x_1 \cdots x_k(1 - x_{k+1})$ fails exactly on the pairs with $\mathbf{x}_i = \mathbf{y}_i = 1$ for all $i \le k$ and $\mathbf{x}_{k+1} \neq \mathbf{y}_{k+1}$, so $\varepsilon(f) = 4^{-k}/2$. Moreover

$\Pr[f = 1 \mid \mathbf{x}_S = 1^S] \le \Pr[\mathbf{x}_{k+1} = 0 \mid \mathbf{x}_S = 1^S] \le \frac{1}{2}$ for every S, so $\Pr[f \ne \mathrm{AND}_S] = \mathbb{E}f + 2^{-|S|}(1 - 2\Pr[f = 1 \mid \mathbf{x}_S = 1^S]) \ge \mathbb{E}f$, and $\mathrm{dist}(f) = \mathbb{E}f = 2^{-k-1} = \sqrt{\varepsilon(f)/2}$. Hence $\delta^*(\varepsilon) \ge \sqrt{\varepsilon/8}$ for all $\varepsilon \le \frac{1}{8}$ (by monotonicity of δ^*), and the substance of the question is the upper bound. The polynomial conjecture is restated, for general predicates, in [2, 9].

1.1 Results

Theorem 1.1 (Main theorem)**.** *Let $f\colon \{0,1\}^n \to \{0,1\}$ with $\mu = \mathbb{E}f > 0$ and $\varepsilon = \varepsilon(f)$, and let $c_0 = 4 + 2\sqrt{2} = 6.828\ldots$ Then*

$$\mathrm{dist}(f) \;\le\; \min\Big\{\mu,\; \frac{c_0\,\varepsilon}{\mu}\Big\} \;\le\; \sqrt{c_0\,\varepsilon}.$$

More precisely, let $\theta_0 = 1 - 1/\sqrt{2}$, $C = \{i \in [n] :\ \Pr[f(\mathbf{x}) = 1,\ \mathbf{x}_i = 0] \le \theta_0\mu\}$ and $\eta = \varepsilon/\mu^2$. If $\eta \le \eta_0 := (2 - \sqrt{2})/4 = 1/c_0$ then

$$\Pr[f \ne \mathrm{AND}_C] \;\le\; h(\eta)\,\frac{\varepsilon}{\mu}, \qquad h(\eta) = (2 + 2\sqrt{2})\,\frac{1 - 2\eta}{1 - (2 + \sqrt{2})\eta},$$

where h increases from $h(0) = 2 + 2\sqrt{2}$ to $h(\eta_0) = c_0$.

A simpler version of the argument (Theorem 5.1 at $p = 1/2$) gives the constant 12 in place of c_0.

Theorem 1.2 (Lower bound)**.** *For every $0 < \varepsilon \le 1/4$ we have $\delta^*(\varepsilon) \ge \sqrt{\varepsilon}$. Moreover, for every $n \ge 1$ the function $f_n = \mathbf{1}\{|x| = n - 1\}$ satisfies $\mathrm{dist}(f_n)^2/\varepsilon(f_n) = n/(n + 1)$.*

Corollary 1.3. *$\sqrt{\varepsilon} \le \delta^*(\varepsilon) \le \sqrt{c_0\,\varepsilon}$ for all $0 < \varepsilon \le 1/4$, where $c_0 = 4 + 2\sqrt{2}$. In particular $\delta^*(\varepsilon) = \Theta(\sqrt{\varepsilon})$: the dependence is polynomial (as conjectured in [11]) with optimal exponent $1/2$.*

The new part of Corollary 1.3 is the upper bound. The order $\sqrt{\varepsilon}$ of the lower bound already follows from $x_1 \cdots x_k(1 - x_{k+1})$ (see above); Theorem 1.2 gives it with the constant 1 for every $\varepsilon \le \frac{1}{4}$, and Section 7 shows that 1 is not the right constant.

For functions close to an AND, the relative form of Theorem 1.1 has a sharp constant. Theorem 5.4 shows that $\Pr[f \ne \mathrm{AND}_C] \le (\frac{1}{2} + O(\sqrt{\eta}))\,\varepsilon/\mu$ for $\eta = \varepsilon/\mu^2 \to 0$, and that $\frac{1}{2}$ cannot be improved. This proof uses small-set expansion; the proof of Theorem 1.1 does not.

The method is robust. We also prove (Section 6):

Theorem 1.4 (Extensions)**.**

(a) (p-biased inputs.) *If $\mathbf{x}, \mathbf{y} \sim \mu_p$ are independent, $p \in (0,1)$, $\mu = \mathbb{E}_{\mu_p} f$ and $\varepsilon = \Pr[f(\mathbf{x} \wedge \mathbf{y}) \ne f(\mathbf{x})f(\mathbf{y})]$, then $\min_{g \in \mathcal{A}_n} \Pr_{\mu_p}[f \ne g] \le \min\{\mu,\ 6\varepsilon/((1-p)\mu)\} \le \sqrt{6\varepsilon/(1-p)}$.*

(b) (m-ary test.) *If $m \ge 2$, $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are independent and uniform and $\varepsilon_m = \Pr[f(\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_m) \ne \prod_j f(\mathbf{x}_j)]$, then $\mathrm{dist}(f) \le \min\{\mu,\ c_0\varepsilon_m/\mu^{m-1}\} \le (c_0\varepsilon_m)^{1/m}$. Conversely, $\delta^*_m(\varepsilon) := \sup_n \sup\{\mathrm{dist}(f) : \varepsilon_m(f) \le \varepsilon\} \ge \varepsilon^{1/m}$ for $0 < \varepsilon < 2^{-m}$. Hence the optimal dependence for the m-ary test is $\Theta(\varepsilon^{1/m})$; more precisely $\varepsilon^{1/m} \le \delta^*_m(\varepsilon) \le (c_0\varepsilon)^{1/m}$.*

(c) (Three functions, two measures; quantitative FLMM Theorem 2.10.) *Let $p, \rho \in (0,1)$, $\mathbf{x} \sim \mu_p$, $\mathbf{y} \sim \mu_\rho$ independent, and $F, G, H\colon \{0,1\}^n \to \{0,1\}$ with $\varepsilon = \Pr[F(\mathbf{x} \wedge \mathbf{y}) \ne G(\mathbf{x})H(\mathbf{y})]$. Put $K = 5/\sqrt{(1-p)(1-\rho)}$. Then either*

(i) $\min(\mathbb{E}_{\mu_p} G, \mathbb{E}_{\mu_\rho} H) \le K\sqrt{\varepsilon}$ *(so* G *or* H *is* $K\sqrt{\varepsilon}$*-close to* 0*) and* $\Pr_{\mu_{p\rho}}[F \neq 0] \le \varepsilon + K\sqrt{\varepsilon}$*; or*

(ii) $\min(\mathbb{E}_{\mu_p} G, \mathbb{E}_{\mu_\rho} H) > K\sqrt{\varepsilon}$*, and there is* $T \subseteq [n]$ *with* $\Pr_{\mu_p}[G \neq \mathrm{AND}_T] \le \frac{6}{5}\sqrt{\frac{1-p}{1-\rho}}\sqrt{\varepsilon}$*,* $\Pr_{\mu_\rho}[H \neq \mathrm{AND}_T] \le \frac{6}{5}\sqrt{\frac{1-\rho}{1-p}}\sqrt{\varepsilon}$*, and* $\Pr_{\mu_{p\rho}}[F \neq \mathrm{AND}_T] \le \varepsilon + \frac{12}{5}\frac{\sqrt{\varepsilon}}{\sqrt{(1-p)(1-\rho)}}$.

Arbitrary predicates. The AND test is the case $g = \mathrm{AND}_2$ of the approximate polymorphism problem of Chase, Filmus, Minzer, Mossel and Saurabh [6] (see also [11, 2]). Given $g\colon \{0,1\}^m \to \{0,1\}$, suppose that $f_0, \dots, f_m\colon \{0,1\}^n \to \{0,1\}$ satisfy

$$f_0\big(g(\mathbf{z}_1), \dots, g(\mathbf{z}_n)\big) = g\big(f_1(\mathbf{Z}^1), \dots, f_m(\mathbf{Z}^m)\big)$$

for most $n \times m$ matrices $\mathbf{Z}$ with rows $\mathbf{z}_i$ and columns $\mathbf{Z}^j$. Are they close to exact solutions? They proved this with a tower-type dependence and asked for the optimal dependence ([6, Open Question 2]). In Section 8 we show that it is always polynomial: a δ-approximate solution is $O_g(\delta^{c_g})$-close to an exact one, with an explicit $c_g > 0$ depending only on the number of relevant coordinates of g (Theorem 8.1). The exponent depends on g (Corollary 8.3): the loss is linear for every balanced g that depends on at least two coordinates (Theorem 8.2, Proposition 8.16), while for every non-constant unbalanced g the optimal dependence is $\delta = \Theta_g(\varepsilon^{e(g)})$, where $e(g) \in \{1, \dots, m\}$ is the number of coordinates of a constant input that must be flipped to change the value of g, maximized over the two constant inputs (Theorem 8.18); for example $e = m$ for $\mathrm{AND}_m, \mathrm{OR}_m, \mathrm{NAND}_m, \mathrm{NOR}_m$. The new ingredient is again the coupling, which yields a junta of size $O(\log(1/\varepsilon))$ with polynomial error in place of the regularity lemma. For functions close to a constant b we prove a lower bound on the error of the form $\Pr[f \neq b]^s$, using a strong data processing inequality for relative entropy. We also determine the optimal dependence for multi-polymorphisms, where $f_0, \dots, f_m$ need not be equal (Theorem 8.28): it is $\delta = \Theta_g(\varepsilon^{E(g)})$, where $E(g) \ge e(g)$ is the largest k such that fixing $m - k$ of the variables of g can leave an AND of k literals or its negation. Hence the loss for multi-polymorphisms is linear if and only if g is affine (Corollary 8.29); for Maj_3 it is linear for polymorphisms but quadratic for multi-polymorphisms. Here some columns may be close to constants while others are not, and we use a partial strong data processing inequality in which only part of the input of the channel is fresh (Lemma 8.30). Finally, Theorem 8.33 extends the polynomial bound to inputs whose rows are i.i.d. samples of an arbitrary full-support distribution μ on $\{0,1\}^m$, the setting of Alekseev and Filmus [2]: every δ-approximate multi-polymorphism is $O_{g,\mu}(\delta^c)$-close to an exact one, with $c = c_{m,\min\mu} > 0$. Independence of the columns is replaced by reverse hypercontractivity [22], and the junta lemma by a latent decomposition of μ. This proves the polynomial part of the conjecture of [2, Section 9] for all functional predicates. Together with [2, Theorem 1.7], its proof also answers the second open question of [2, Section 9] for Boolean functional predicates (Remark 8.44). Over a finite alphabet Σ (Theorem 8.45) we prove polynomial dependence for every $g\colon \Sigma^m \to \Sigma$ such that changes of a single coordinate x_j within a connected component of its collision graph ($a \sim b$ if $g(u^{j\to a}) = g(u^{j\to b})$ for some context u) connect all output values; this holds, for instance, when every coordinate is irrelevant under some context. Examples are all finite semilattice operations (e.g. min and max), all near-unanimity operations (e.g. majority on Σ^3), products in finite monoids with zero (e.g. multiplication modulo q), and the rounded mean $\lfloor (x_1+x_2)/2 \rfloor$. A new ingredient is a junta lemma in which column j is resampled only within one edge of the collision graph at a time; an alternating-projection estimate (Lemma 8.49) combines these partial bounds, and f_0 is reached by resampling each row within its vector of components. The flexibility condition of [2, Theorem 1.10] fails for every functional predicate, so even the qualitative statement for these predicates is not covered by [2]. Alekseev and Filmus name

$P_{\not=} = \{(a,b,c) \in \mathbb{Z}_3^3 : a+b+c \neq 0\}$ as the simplest predicate that their techniques cannot handle. Under the uniform distribution on $P_{\not=}$ we prove that δ-approximate generalized polymorphisms are $(1+o(1))\sqrt{\delta}$-close to exact ones, and that this is sharp, including the constant (Theorem 8.55); the dependence is linear near the affine dictator family. (The classification of the exact ones, the case $\delta = 0$, also follows from a theorem of Filmus [10] and a finite check.) Here almost every restriction to a single coordinate is an exact solution in one variable, and pairwise independence of the columns removes the three-wise Fourier terms. For every full-support distribution μ on $P_{\not=}$ with uniform marginals (an 11-dimensional family containing all translation-invariant distributions), the distance is $O(\alpha^{-3/2}\delta^{4\alpha})$ with $\alpha = \min \mu$ (Theorem 8.61): a latent decomposition $\mu = 9\alpha\,\pi + (1-9\alpha)\nu$ lets us apply Theorem 8.55 to restrictions, and reverse hypercontractivity and Harris's inequality glue the restrictions together. If only two marginals are uniform (a 13-dimensional family), the distance is $O(\alpha^{-3/2}\delta^{32\alpha^2/9})$. The exponent must depend on μ: for a family $\mu^{(\alpha)}$ with uniform marginals and $\min \mu^{(\alpha)} = \alpha$ it is at most $\log 3 / \log \frac{1}{2\alpha}$, and for another such family, μ_α^{sh}, at most 9α (Propositions 8.62 and 8.82); for every full-support μ with uniform marginals it is at most $\frac{9}{4}(\frac{1}{3} - M)$, where M is the largest probability of a pair of values of two entries of a row. With a smaller weight of π in the decomposition, every exponent below $r_{18\alpha}/2 \geq \max(8\alpha, 9\alpha(1-9\alpha))$ is valid, where $r_\kappa = 2(1-\sqrt{1-\kappa})/(2-\sqrt{1-\kappa}) \in [\frac{8}{9}\kappa, \kappa)$ is the exponent of reverse hypercontractivity in Lemma 8.35, and every exponent below $32\alpha^2$ if only two marginals are uniform (Proposition 8.81). So for uniform marginals and smallest atom $\alpha < \frac{1}{18}$, the best exponent valid for all such μ lies in $[r_{18\alpha}/2, 9\alpha]$, an interval whose ends are within a factor $\frac{9}{8}$; both ends tend to the optimal exponent $\frac{1}{2}$ of π as $\alpha \to \frac{1}{18}$. The factor $\frac{9}{8}$ is exactly the loss in reverse hypercontractivity, and we remove it for $\alpha \leq \frac{1}{36}$: when $\kappa \leq \frac{1}{2}$, the simple operator $(1-\kappa)\mathrm{Id} + \kappa\mathbb{E}$ on every probability space satisfies the two-function reverse hypercontractive inequality with the sharp exponent κ for both functions (the analogue of Borell's two-point inequality), so for $\alpha \leq \frac{1}{36}$ the best exponent is exactly 9α; for $\frac{1}{36} < \alpha < \frac{1}{18}$ it would be 9α if the same inequality holds on the uniform three-point space for $\kappa \in (\frac{1}{2}, 1)$, which we conjecture (Proposition 8.83). Near the affine family the dependence is linear for every full-support μ (Lemma 8.60). Finally, every full-support μ is covered (Theorem 8.66): the distance is at most $C_\alpha \delta^{\alpha^8/13107200}$. A stable distribution π' that shares two marginals with μ can replace π in the decomposition (Lemma 8.64), and three such steps (Lemma 8.65) connect every μ to a distribution with two uniform marginals. This answers the first question of [2, Section 9] for $P_{\not=}$. The guarantee of their second question (asked there for Boolean predicates), that equal functions with equal marginals get equal approximations, holds for $P_{\not=}$ as well, and for polymorphisms, i.e. $f_1 = f_2 = f_3$, the dependence is linear whenever the three marginals coincide; under π the sharp constant is $\frac{2}{3}$ (Corollary 8.68). The optimal exponent $\frac{1}{2}$ and sharp constant 1 of Theorem 8.55 also hold for $\{a \in \mathbb{Z}_3^m : a_1 + \cdots + a_m \neq 0\}$, $m \geq 4$, under the uniform distribution, with distance at most $\sqrt{\delta} + O(m^2\delta)$ (Theorem 8.70): conditioning on all but three columns reduces the case of two functions far from constants to Theorem 8.55, and the case of nearly constant functions is handled directly. For their polymorphisms the dependence is linear with sharp constant $\frac{2}{m}$ if $3 \mid m$, and a square root with sharp constant $\sqrt{2/(m(m-1))}$ if $3 \nmid m$, where the constants 1 and 2 are polymorphisms (Corollary 8.75). For $m \geq 3$ these predicates are, up to relabelling the alphabet in each coordinate, the only predicates over $\mathbb{Z}_3$ of the largest possible size $2 \cdot 3^{m-1}$ without a flexible coordinate (Corollary 8.73; this follows from the classical fact that all multiary quasigroups (Latin hypercubes) of order 3 are affine [20, 27, 19]). So all of them are covered, under every full-support distribution μ. For $m \geq 4$ (Theorem 8.78), each pair of coordinates b, c and a third coordinate a, μ is a mixture of a remainder and the distribution under which $T(y) = (y_a + \sum_{j \in R} y_j, y_b, y_c)$ and $(y_j)_{j \in R}$ are independent with their laws under μ, where $R = [m] \setminus \{a,b,c\}$; the law of $T(y)$ is a full-support distribution on $P_{\not=}$, so it is stable by the result for $m = 3$, and the transfer argument

of Lemma 8.64 shows that f_b or f_c is close to a constant. The exponent must again depend on μ. For uniform marginals the distance is $O_{m,\alpha}(\delta^{4\alpha'})$ with $\alpha' = 3^{m-3}\alpha$ (Theorem 8.80); every exponent below $r_{18\alpha'}/2$ is valid (Proposition 8.81), and for some such μ no exponent above $9\alpha'$ is possible (Proposition 8.82), so again, for $\alpha' < \frac{1}{18}$, the best exponent lies in $[r_{18\alpha'}/2, 9\alpha']$, uniformly in m, and it equals $9\alpha'$ for $\alpha' \le \frac{1}{36}$ (Proposition 8.83).

The idea. FLMM's approach passes through the approximate eigenvalue equation $\mathrm{T}^{\downarrow} f \approx \mu f$, where $\mathrm{T}^{\downarrow} f(x) = \mathbb{E}_{\mathbf{y}} f(x \wedge \mathbf{y})$; this only uses the AND test averaged over $\mathbf{y}$, it needs a junta theorem, and the junta must then be made exact, which is where the logarithmic loss comes from. We use the test pointwise through a simple coupling: re-randomize $\mathbf{x}$ on the zero set of $\mathbf{y}$. This does not change $\mathbf{x} \wedge \mathbf{y}$, so if $f(\mathbf{y}) = 1$ and f changes, the test must fail for one of the two pairs. Averaging over $\mathbf{y}$ turns this into a single inequality (Lemma 3.1) that controls the *whole* Fourier spectrum of f at once:

$$\sum_{S \subseteq [n]} \hat{f}(S)^2 \Pr\left[f(\mathbf{x}) = 1,\ \mathbf{x}_S \neq 1^S\right] \ \le\ \varepsilon.$$

Normalized, it says that the spectral sample of f is almost always contained in the support of a random point of $f^{-1}(1)$. Combined with an elementary lower bound on the spectral weight of "core" coordinates (Lemma 2.2), this forces $f^{-1}(1)$ to sit inside the subcube $\{x_C = 1^C\}$ and f to be essentially a function of x_C, hence close to AND_C (Section 4). The constant c_0 comes from a slightly finer accounting in this last step (Lemma 4.3).

The constant. Slices and thresholds approach $\mathrm{dist}(f)^2/\varepsilon(f) = 1$ from below (Theorem 1.2), and exhaustive computation over all 2^{16} functions on 4 variables and over all symmetric functions on $n \le 16$ variables never exceeds 1. Nevertheless the constant 1 is not sharp. In Section 7 we exhibit a symmetric function on 69 variables with $\mathrm{dist}(f)^2 > 1.12\,\varepsilon(f)$, and a symmetric function on 16000 variables with $\mathrm{dist}(f)^2 > 1.342\,\varepsilon(f)$ (both by exact integer computation); a padding lemma transfers them to arbitrarily small ε. As $n \to \infty$ these two-band constructions have a Gaussian limit, which we compute rigorously (central limit theorem and a ball-arithmetic evaluation of a one-dimensional integral; Proposition 7.4): there are functions with $\mathrm{dist}(f)^2 > 1.351\,\varepsilon(f)$ at arbitrarily small ε. So $1.1623 < \limsup_{\varepsilon \to 0} \delta^*(\varepsilon)/\sqrt{\varepsilon} \le \sqrt{c_0} < 2.614$.

2 Preliminaries

Throughout, $p \in (0,1)$ and μ_p denotes the p-biased product measure on $\{0,1\}^n$, $\mu_p(x) = p^{|x|}(1-p)^{n-|x|}$; $\mu_{1/2}$ is the uniform measure. For $x \in \{0,1\}^n$ let $\mathrm{supp}(x) = \{i : x_i = 1\}$ and $Z(x) = [n] \setminus \mathrm{supp}(x)$. For $J \subseteq [n]$, x_J denotes the restriction of x to J, and $x_J = 1^J$ means $x_i = 1$ for all $i \in J$ (always true for $J = \emptyset$).

Fourier analysis. For $S \subseteq [n]$ let $\phi_S(x) = \prod_{i \in S} \frac{x_i - p}{\sqrt{p(1-p)}}$. The ϕ_S form an orthonormal basis of $L^2(\{0,1\}^n, \mu_p)$; we write $\hat{f}(S) = \mathbb{E}_{\mu_p}[f\phi_S]$ (the measure will always be clear from the context). Parseval: $\mathbb{E} f^2 = \sum_S \hat{f}(S)^2$. For $J \subseteq [n]$ and $\mathbf{x} \sim \mu_p$, the conditional expectation is

$$\mathbb{E}[f(\mathbf{x}) \mid \mathbf{x}_J] = \sum_{S \subseteq J} \hat{f}(S)\phi_S(\mathbf{x}), \qquad \text{hence} \qquad \mathbb{E}\left[(f - \mathbb{E}[f \mid \mathbf{x}_J])^2\right] = \sum_{S \not\subseteq J} \hat{f}(S)^2. \tag{1}$$

(Indeed $\mathbb{E}[\phi_S(\mathbf{x}) \mid \mathbf{x}_J] = \phi_S(\mathbf{x})$ if $S \subseteq J$ and $= 0$ otherwise, since the factors $\frac{\mathbf{x}_i - p}{\sqrt{p(1-p)}}$, $i \notin J$, are independent of $\mathbf{x}_J$ and have mean 0.)

Notation for a Boolean function. For $f\colon \{0,1\}^n \to \{0,1\}$ on $(\{0,1\}^n, \mu_p)$ with $\mu = \mathbb{E}f > 0$ write $A = f^{-1}(1)$ and let $P_A = \mu_p(\,\cdot \mid A)$. Since $f^2 = f$, Parseval gives $\sum_S \hat{f}(S)^2 = \mu$, so

$$w_S^f = \hat{f}(S)^2/\mu \qquad (S \subseteq [n])$$

is a probability distribution on subsets of $[n]$ (the *spectral distribution* of f). Note that $\Pr_{\mu_p}[f = 1,\ \mathbf{x}_S \neq 1^S] = \mu\, P_A[\mathbf{x}_S \neq 1^S]$.

Lemma 2.1 (Resampling)**.** *Let $f\colon \{0,1\}^n \to \{0,1\}$, $J \subseteq [n]$, $\mathbf{x} \sim \mu_p$, and let $\mathbf{x}'$ agree with $\mathbf{x}$ on J while $\mathbf{x}'_{[n]\setminus J}$ is an independent fresh sample of $\mu_p^{[n]\setminus J}$. Then $\mathbf{x}' \sim \mu_p$ and*

$$\Pr[f(\mathbf{x}) \neq f(\mathbf{x}')] = 2 \sum_{S \not\subseteq J} \hat{f}(S)^2.$$

Proof. Clearly $\mathbf{x}' \sim \mu_p$. For $S, T \subseteq [n]$ we have $\mathbb{E}[\phi_S(\mathbf{x})\phi_T(\mathbf{x}')] = \prod_i \mathbb{E}[\cdots]$, where for $i \in J$ the factor is 1 if $[i \in S] = [i \in T]$ and 0 otherwise (orthonormality in one coordinate), and for $i \notin J$ the factor is 1 if $i \notin S \cup T$ and 0 otherwise (independence and mean zero). Hence $\mathbb{E}[\phi_S(\mathbf{x})\phi_T(\mathbf{x}')] = \mathbf{1}\{S = T \subseteq J\}$ and $\mathbb{E}[f(\mathbf{x})f(\mathbf{x}')] = \sum_{S \subseteq J} \hat{f}(S)^2$. Since f is $\{0,1\}$-valued, $\Pr[f(\mathbf{x}) \neq f(\mathbf{x}')] = \mathbb{E}f(\mathbf{x}) + \mathbb{E}f(\mathbf{x}') - 2\mathbb{E}[f(\mathbf{x})f(\mathbf{x}')] = 2\mu - 2\sum_{S \subseteq J} \hat{f}(S)^2 = 2\sum_{S \not\subseteq J} \hat{f}(S)^2$. □

Lemma 2.2 (Spectral weight of a coordinate)**.** *Let $f\colon (\{0,1\}^n, \mu_p) \to \{0,1\}$, $\mu = \mathbb{E}f$, $i \in [n]$ and $b_i = \Pr[f(\mathbf{x}) = 1,\ \mathbf{x}_i = 0]$. Then*

$$\sum_{S \ni i} \hat{f}(S)^2 \ \geq\ (1-p)\mu - b_i.$$

Equivalently, if $\mu > 0$: $\Pr_{\mathbf{S} \sim w^f}[i \in \mathbf{S}] \geq (1-p) - P_A[\mathbf{x}_i = 0]$.

Proof. Let $f^b(x) = f(x_1, \ldots, x_{i-1}, b, x_{i+1}, \ldots, x_n)$ for $b \in \{0,1\}$ and $E_i f = p f^1 + (1-p) f^0$. By (1) with $J = [n] \setminus \{i\}$, $\sum_{S \ni i} \hat{f}(S)^2 = \mathbb{E}[(f - E_i f)^2]$. If $\mathbf{x}_i = 1$ then $f - E_i f = (1-p)(f^1 - f^0)$, and if $\mathbf{x}_i = 0$ then $f - E_i f = -p(f^1 - f^0)$; as $f^1 - f^0$ does not depend on $\mathbf{x}_i$,

$$\mathbb{E}[(f - E_i f)^2] = \big(p(1-p)^2 + (1-p)p^2\big)\mathbb{E}[(f^1 - f^0)^2] = p(1-p)\,\mathbb{E}|f^1 - f^0| \ \geq\ p(1-p)\big(\mathbb{E}f^1 - \mathbb{E}f^0\big),$$

using that $f^1 - f^0 \in \{-1, 0, 1\}$. Finally $\mathbb{E}f^1 = \Pr[f = 1 \mid \mathbf{x}_i = 1] = (\mu - b_i)/p$ and $\mathbb{E}f^0 = b_i/(1-p)$, so the right-hand side equals $(1-p)(\mu - b_i) - p b_i = (1-p)\mu - b_i$. □

3 The spectral containment inequality

Lemma 3.1 (Spectral containment, three functions)**.** *Let $p, \rho \in (0,1)$, let $\mathbf{x} \sim \mu_p$ and $\mathbf{y} \sim \mu_\rho$ be independent, and let $F, G, H\colon \{0,1\}^n \to \{0,1\}$. Let $\varepsilon = \Pr[F(\mathbf{x} \wedge \mathbf{y}) \neq G(\mathbf{x})H(\mathbf{y})]$. Then, with $\hat{G}$ the μ_p-Fourier coefficients of G,*

$$\sum_{S \subseteq [n]} \hat{G}(S)^2 \Pr_{\mathbf{y} \sim \mu_\rho}\big[H(\mathbf{y}) = 1,\ \mathbf{y}_S \neq 1^S\big] \ \leq\ \varepsilon.$$

Proof. Let $\mathbf{w} \sim \mu_p$ be independent of $(\mathbf{x}, \mathbf{y})$ and define $\mathbf{x}'_i = \mathbf{x}_i$ if $\mathbf{y}_i = 1$ and $\mathbf{x}'_i = \mathbf{w}_i$ if $\mathbf{y}_i = 0$. Conditionally on $\mathbf{y}$, the coordinates $\mathbf{x}'_1, \ldots, \mathbf{x}'_n$ are independent Bernoulli(p) variables, so $\mathbf{x}' \sim \mu_p$ independently of $\mathbf{y}$, and $(\mathbf{x}', \mathbf{y})$ has the same distribution as $(\mathbf{x}, \mathbf{y})$. Moreover $\mathbf{x} \wedge \mathbf{y} = \mathbf{x}' \wedge \mathbf{y}$, because the two points agree on $\mathrm{supp}(\mathbf{y})$ and both meets vanish on $Z(\mathbf{y})$. Finally, conditionally on $\mathbf{y}$, the pair $(\mathbf{x}, \mathbf{x}')$ is exactly the pair of Lemma 2.1 with $J = \mathrm{supp}(\mathbf{y})$.

Consider the event $E = \{H(\mathbf{y}) = 1,\ G(\mathbf{x}) \neq G(\mathbf{x}')\}$. On E we have $G(\mathbf{x})H(\mathbf{y}) = G(\mathbf{x}) \neq G(\mathbf{x}') = G(\mathbf{x}')H(\mathbf{y})$, whereas $F(\mathbf{x} \wedge \mathbf{y}) = F(\mathbf{x}' \wedge \mathbf{y})$; hence $F(\mathbf{x} \wedge \mathbf{y}) \neq G(\mathbf{x})H(\mathbf{y})$ or $F(\mathbf{x}' \wedge \mathbf{y}) \neq G(\mathbf{x}')H(\mathbf{y})$. As both pairs are distributed like $(\mathbf{x}, \mathbf{y})$, $\Pr[E] \leq 2\varepsilon$.

On the other hand, conditioning on $\mathbf{y}$ and applying Lemma 2.1 (to G with $J = \mathrm{supp}(\mathbf{y})$),

$$\Pr[E] = \mathbb{E}_{\mathbf{y}}\Big[H(\mathbf{y}) \cdot 2 \sum_{S \not\subseteq \mathrm{supp}(\mathbf{y})} \hat{G}(S)^2\Big] = 2 \sum_{S} \hat{G}(S)^2 \Pr_{\mathbf{y}}[H(\mathbf{y}) = 1,\ S \not\subseteq \mathrm{supp}(\mathbf{y})].$$

Since $S \not\subseteq \mathrm{supp}(\mathbf{y})$ iff $\mathbf{y}_S \neq 1^S$, the lemma follows. □

Corollary 3.2 (One function)**.** *Let $f\colon \{0,1\}^n \to \{0,1\}$, $\mathbf{x}, \mathbf{y} \sim \mu_p$ independent, $\mu = \mathbb{E}f > 0$, $\varepsilon = \Pr[f(\mathbf{x} \wedge \mathbf{y}) \neq f(\mathbf{x})f(\mathbf{y})]$. Then*

$$\sum_S \hat{f}(S)^2\big(\mu - \Pr[f(\mathbf{x}) = 1,\ \mathbf{x}_S = 1^S]\big) \leq \varepsilon, \qquad \textit{i.e.} \qquad \Pr_{\mathbf{S} \sim w^f,\ \mathbf{x} \sim P_A}\big[\mathbf{S} \not\subseteq \mathrm{supp}(\mathbf{x})\big] \leq \frac{\varepsilon}{\mu^2}.$$

Proof. Apply Lemma 3.1 with $\rho = p$ and $F = G = H = f$, and note $\Pr[f = 1, \mathbf{x}_S \neq 1^S] = \mu - \Pr[f = 1, \mathbf{x}_S = 1^S] = \mu P_A[\mathbf{x}_S \neq 1^S]$; divide by μ^2. □

Corollary 3.3 (m-ary test)**.** *Let $m \geq 2$, $\mathbf{x}_1, \ldots, \mathbf{x}_m \sim \mu_p$ independent, $\mu = \mathbb{E}f > 0$ and $\varepsilon_m = \Pr[f(\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_m) \neq \prod_j f(\mathbf{x}_j)]$. Then $\Pr_{\mathbf{S} \sim w^f,\ \mathbf{x} \sim P_A}[\mathbf{S} \not\subseteq \mathrm{supp}(\mathbf{x})] \leq \varepsilon_m/\mu^m$.*

Proof. Let $\mathbf{z} = \mathbf{x}_2 \wedge \cdots \wedge \mathbf{x}_m$ and let $\mathbf{x}_1'$ agree with $\mathbf{x}_1$ on $\mathrm{supp}(\mathbf{z})$ and be fresh on $Z(\mathbf{z})$. Then $(\mathbf{x}_1', \mathbf{x}_2, \ldots, \mathbf{x}_m)$ has the law of $(\mathbf{x}_1, \ldots, \mathbf{x}_m)$, and the two m-tuples have the same meet. If $f(\mathbf{x}_2) = \cdots = f(\mathbf{x}_m) = 1$ and $f(\mathbf{x}_1) \neq f(\mathbf{x}_1')$, the test fails on one of the two tuples. Lemma 2.1 (given $\mathbf{x}_2, \ldots, \mathbf{x}_m$, with $J = \mathrm{supp}(\mathbf{z})$) yields

$$2 \sum_S \hat{f}(S)^2 \Pr\big[f(\mathbf{x}_2) = \cdots = f(\mathbf{x}_m) = 1,\ S \not\subseteq \mathrm{supp}(\mathbf{z})\big] \leq 2\varepsilon_m.$$

Since $\mathrm{supp}(\mathbf{z}) \subseteq \mathrm{supp}(\mathbf{x}_2)$, the probability is at least $\Pr[f(\mathbf{x}_2) = \cdots = f(\mathbf{x}_m) = 1,\ S \not\subseteq \mathrm{supp}(\mathbf{x}_2)] = \mu^{m-2} \cdot \mu P_A[\mathbf{x}_S \neq 1^S]$. Dividing by μ^m gives the claim. □

4 From spectral containment to an AND

For Boolean G on $(\{0,1\}^n, \mu_p)$ and H on $(\{0,1\}^n, \mu_\rho)$, both not identically 0, define

$$\eta_{G|H} = \sum_S w_S^G\, P_H[\mathbf{y}_S \neq 1^S] = \Pr_{\mathbf{S} \sim w^G,\ \mathbf{y} \sim P_H}\big[\mathbf{S} \not\subseteq \mathrm{supp}(\mathbf{y})\big],$$

where $P_H = \mu_\rho(\cdot \mid H = 1)$. Lemma 3.1 says exactly that $\eta_{G|H} \leq \varepsilon/(\mathbb{E}G\,\mathbb{E}H)$, and by symmetry ($\mathbf{x} \wedge \mathbf{y} = \mathbf{y} \wedge \mathbf{x}$) also $\eta_{H|G} \leq \varepsilon/(\mathbb{E}G\,\mathbb{E}H)$.

Lemma 4.1 (Final step)**.** *Let $G\colon (\{0,1\}^n, \mu_p) \to \{0,1\}$ with $\mu_G = \mathbb{E}G > 0$, let $C \subseteq [n]$ and $0 \leq \kappa < 1/2$. Suppose*

$$\textit{(a)}\ \sum_{S \not\subseteq C} \hat{G}(S)^2 \leq \kappa\mu_G \qquad \textit{and} \qquad \textit{(b)}\ \Pr[G(\mathbf{x}) = 1,\ \mathbf{x}_C \neq 1^C] \leq \kappa\mu_G.$$

Then $\Pr[G \neq \mathrm{AND}_C] \leq \kappa\mu_G \frac{2-3\kappa}{1-2\kappa}$. In particular $\Pr[G \neq \mathrm{AND}_C] \leq 3\kappa\mu_G$ if $\kappa \leq 1/3$.

Proof. Let $q = \Pr[\mathbf{x}_C = 1^C]$ and $m_1 = \mathbb{E}[G \mid \mathbf{x}_C = 1^C]$. By (b), $qm_1 = \Pr[G = 1, \mathbf{x}_C = 1^C] = \mu_G - \Pr[G = 1, \mathbf{x}_C \neq 1^C] \geq (1-\kappa)\mu_G > 0$. By (1) and (a),

$$\kappa\mu_G \;\geq\; \mathbb{E}\big[(G - \mathbb{E}[G \mid \mathbf{x}_C])^2\big] = \sum_{u \in \{0,1\}^C} \Pr[\mathbf{x}_C = u]\,\mathrm{Var}(G \mid \mathbf{x}_C = u) \;\geq\; q\,m_1(1-m_1),$$

because G is Boolean. Hence $1 - m_1 \leq \kappa\mu_G/(qm_1) \leq \kappa/(1-\kappa)$, i.e. $m_1 \geq (1-2\kappa)/(1-\kappa) > 0$, and

$$\Pr[G = 0,\ \mathbf{x}_C = 1^C] = q(1-m_1) = \frac{q\,m_1(1-m_1)}{m_1} \leq \frac{\kappa\mu_G(1-\kappa)}{1-2\kappa}.$$

Adding (b): $\Pr[G \neq \mathrm{AND}_C] = \Pr[G = 1, \mathbf{x}_C \neq 1^C] + \Pr[G = 0, \mathbf{x}_C = 1^C] \leq \kappa\mu_G + \frac{\kappa\mu_G(1-\kappa)}{1-2\kappa} = \kappa\mu_G\frac{2-3\kappa}{1-2\kappa}$. The function $\kappa \mapsto \frac{2-3\kappa}{1-2\kappa}$ is increasing on $[0, 1/2)$ (its derivative is $(1-2\kappa)^{-2}$) and equals 3 at $\kappa = 1/3$. □

Proposition 4.2 (Structure)**.** *Let $G\colon (\{0,1\}^n, \mu_p) \to \{0,1\}$ and $H\colon (\{0,1\}^n, \mu_\rho) \to \{0,1\}$ be not identically zero, let $0 < \theta < 1 - \rho$, and*

$$C = C_\theta(H) = \{i \in [n] :\ P_H[\mathbf{y}_i = 0] \leq \theta\}.$$

Then

(a) $\Pr_{\mathbf{S}\sim w^G}[\mathbf{S} \not\subseteq C] \leq \eta_{G|H}/\theta$, *i.e.* $\sum_{S \not\subseteq C} \hat{G}(S)^2 \leq \mu_G\,\eta_{G|H}/\theta$;

(b) $P_G[\mathbf{x}_C \neq 1^C] \leq \eta_{H|G}/(1-\rho-\theta)$, *i.e.* $\Pr_{\mu_p}[G = 1,\ \mathbf{x}_C \neq 1^C] \leq \mu_G\,\eta_{H|G}/(1-\rho-\theta)$.

Consequently, if $\kappa := \max\{\eta_{G|H}/\theta,\ \eta_{H|G}/(1-\rho-\theta)\} < 1/2$, then $\Pr_{\mu_p}[G \neq \mathrm{AND}_C] \leq \kappa\mu_G\frac{2-3\kappa}{1-2\kappa}$.

Proof. (a) If $S \not\subseteq C$, pick $i \in S \setminus C$; then $P_H[\mathbf{y}_S \neq 1^S] \geq P_H[\mathbf{y}_i = 0] > \theta$. Hence $\eta_{G|H} \geq \sum_{S \not\subseteq C} w_S^G P_H[\mathbf{y}_S \neq 1^S] \geq \theta \sum_{S \not\subseteq C} w_S^G$.

(b) For a set $Z \subseteq [n]$, $\Pr_{\mathbf{S}\sim w^H}[\mathbf{S} \cap Z \neq \emptyset] \geq \Pr_{\mathbf{S}\sim w^H}[i \in \mathbf{S}]$ for every $i \in Z$. If $i \in C$, Lemma 2.2 (applied to H on $(\{0,1\}^n, \mu_\rho)$) gives $\Pr_{w^H}[i \in \mathbf{S}] \geq (1-\rho) - P_H[\mathbf{y}_i = 0] \geq 1 - \rho - \theta$. Therefore, writing $\eta_{H|G} = \mathbb{E}_{\mathbf{x}\sim P_G} \Pr_{\mathbf{S}\sim w^H}[\mathbf{S} \cap Z(\mathbf{x}) \neq \emptyset]$,

$$\eta_{H|G} \;\geq\; \mathbb{E}_{\mathbf{x}\sim P_G}\Big[\mathbf{1}\{Z(\mathbf{x}) \cap C \neq \emptyset\}\,(1-\rho-\theta)\Big] = (1-\rho-\theta)\,P_G[\mathbf{x}_C \neq 1^C].$$

The last assertion is Lemma 4.1 applied with (a) and (b). □

Lemma 4.1 does not use the definition of C. On the uniform cube, using it (every coordinate of C is rarely 0 on $G^{-1}(1)$) improves the constants.

Lemma 4.3 (Final step, refined)**.** *Let $G\colon \{0,1\}^n \to \{0,1\}$ with $\mu_G = \mathbb{E}G > 0$ under the uniform measure $\mu_{1/2}$, let $C \subseteq [n]$, $0 < \theta < 1$ and $0 \leq \kappa_a, \kappa_b < 1$. Suppose*

$$(a)\ \sum_{S \not\subseteq C} \hat{G}(S)^2 \leq \kappa_a\mu_G, \qquad (b)\ \Pr[G = 1,\ \mathbf{x}_C \neq 1^C] \leq \kappa_b\mu_G,$$
$$(c)\ \Pr[G = 1,\ \mathbf{x}_i = 0] \leq \theta\mu_G \ \text{ for every } i \in C.$$

Then

$$\frac{\Pr[G \neq \mathrm{AND}_C]}{\mu_G} \;\leq\; \max\Big\{\frac{\kappa_a}{1-\kappa_a},\ \ 2\kappa_b - 1 + \frac{(1-\kappa_b)^2 + \theta\kappa_b}{1-\kappa_a},\ \ 2\kappa_b - 1 + \theta\Big\}.$$

Proof. Let $q = 2^{-|C|} = \Pr[\mathbf{x}_C = u]$ for every $u \in \{0,1\}^C$, let $m_u = \mathbb{E}[G \mid \mathbf{x}_C = u]$ and $m = m_{1^C}$. Put $\beta = \Pr[G = 1, \mathbf{x}_C \neq 1^C]/\mu_G \in [0, \kappa_b]$ and $s = 1 - \beta \in [1 - \kappa_b, 1]$; note $s > 0$.

(i) $qm = \Pr[G = 1, \mathbf{x}_C = 1^C] = \mu_G - \beta\mu_G = s\mu_G > 0$; in particular $m > 0$.

(ii) Let $u \neq 1^C$ and pick $i \in C$ with $u_i = 0$. By (c) and (i), $qm_u = \Pr[G = 1, \mathbf{x}_C = u] \leq \Pr[G = 1, \mathbf{x}_i = 0] \leq \theta\mu_G = \theta qm/s$, so $m_u \leq \theta m/s$.

(iii) By (1) and (a), and since $\mathrm{Var}(G \mid \mathbf{x}_C = u) = m_u(1 - m_u)$ for Boolean G,

$$\kappa_a \mu_G \;\geq\; \mathbb{E}\big[(G - \mathbb{E}[G \mid \mathbf{x}_C])^2\big] = \sum_u q\, m_u(1 - m_u).$$

The term $u = 1^C$ equals $s\mu_G(1 - m)$ by (i). For $u \neq 1^C$ we have $1 - m_u \geq (1 - \theta m/s)_+$ by (ii), where $t_+ = \max(t, 0)$, and $\sum_{u \neq 1^C} qm_u = \Pr[G = 1, \mathbf{x}_C \neq 1^C] = (1 - s)\mu_G$. Therefore

$$\kappa_a \;\geq\; s(1 - m) + (1 - s)\Big(1 - \frac{\theta m}{s}\Big)_+. \tag{2}$$

(iv) By (i), $q = s\mu_G/m$. Since $\Pr[G \neq \mathrm{AND}_C] = \Pr[G = 1, \mathbf{x}_C \neq 1^C] + \Pr[G = 0, \mathbf{x}_C = 1^C]$,

$$\Pr[G \neq \mathrm{AND}_C] = (1 - s)\mu_G + q(1 - m) = \mu_G\Big(1 - 2s + \frac{s}{m}\Big).$$

Case $\theta m \leq s$. Then (2) reads $\kappa_a \geq s - sm + (1-s) - \theta m(1-s)/s$, i.e. $\frac{m}{s}\big(s^2 + \theta(1-s)\big) \geq 1 - \kappa_a > 0$. Hence $s/m \leq (s^2 + \theta(1 - s))/(1 - \kappa_a)$ and, by (iv),

$$\frac{\Pr[G \neq \mathrm{AND}_C]}{\mu_G} \;\leq\; B(s) := 1 - 2s + \frac{s^2 + \theta(1 - s)}{1 - \kappa_a}.$$

B is a convex quadratic function of s, so on the interval $[1 - \kappa_b, 1]$ it is maximised at an endpoint; $B(1) = \kappa_a/(1 - \kappa_a)$ and $B(1 - \kappa_b) = 2\kappa_b - 1 + \frac{(1-\kappa_b)^2 + \theta\kappa_b}{1 - \kappa_a}$.

Case $\theta m > s$. Then $s/m < \theta$ and, by (iv), $\Pr[G \neq \mathrm{AND}_C]/\mu_G < 1 - 2s + \theta \leq 2\kappa_b - 1 + \theta$. □

5 Proofs of the main theorem and of the lower bound

Theorem 5.1 (Main theorem, p-biased form)**.** *Let $p \in (0,1)$, $f\colon \{0,1\}^n \to \{0,1\}$, $\mathbf{x}, \mathbf{y} \sim \mu_p$ independent, $\mu = \mathbb{E}_{\mu_p} f > 0$, $\varepsilon = \Pr[f(\mathbf{x} \wedge \mathbf{y}) \neq f(\mathbf{x})f(\mathbf{y})]$. Let $C = \{i : \Pr_{\mu_p}[f = 1, \mathbf{x}_i = 0] \leq \frac{1-p}{2}\mu\}$ and $\kappa = \frac{2\varepsilon}{(1-p)\mu^2}$. If $\kappa < 1/2$ then $\Pr_{\mu_p}[f \neq \mathrm{AND}_C] \leq \kappa\mu\frac{2-3\kappa}{1-2\kappa}$, and in all cases*

$$\min_{g \in \mathcal{A}_n} \Pr_{\mu_p}[f \neq g] \;\leq\; \min\Big\{\mu,\; \frac{6\varepsilon}{(1-p)\mu}\Big\} \;\leq\; \sqrt{\frac{6\varepsilon}{1-p}}.$$

Proof. Apply Proposition 4.2 with $G = H = f$, $\rho = p$ and $\theta = (1 - p)/2$; then $C_\theta(f)$ is the set C of the statement, and by Corollary 3.2, $\eta_{f|f} \leq \varepsilon/\mu^2$, so the κ of the proposition is at most $\frac{2\varepsilon}{(1-p)\mu^2}$. This gives the first claim (the bound is increasing in κ). For the second, if $\kappa \leq 1/3$ then $\Pr[f \neq \mathrm{AND}_C] \leq 3\kappa\mu = \frac{6\varepsilon}{(1-p)\mu}$, while if $\kappa > 1/3$ then $\frac{6\varepsilon}{(1-p)\mu} = 3\kappa\mu > \mu = \Pr[f \neq 0]$. Finally $\min\{\mu, a/\mu\} \leq \sqrt{a}$ for all $a, \mu > 0$. □

At $p = 1/2$ Theorem 5.1 gives $\mathrm{dist}(f) \leq \min\{\mu, 12\varepsilon/\mu\}$. The constant c_0 of Theorem 1.1 needs Lemma 4.3.

Lemma 5.2 (Core bound). *Let $f\colon \{0,1\}^n \to \{0,1\}$ with $\mu = \mathbb{E}f > 0$ (uniform measure), and let η satisfy $\eta_{f|f} \le \eta \le \eta_0 = (2-\sqrt{2})/4$. Let $\theta_0 = 1 - 1/\sqrt{2}$ and $C = C_{\theta_0}(f) = \{i : \Pr[f = 1, \mathbf{x}_i = 0] \le \theta_0\mu\}$. Then*

$$\Pr[f \ne \mathrm{AND}_C] \;\le\; h(\eta)\,\eta\,\mu \;\le\; c_0\,\eta\,\mu, \qquad h(\eta) = (2+2\sqrt{2})\frac{1-2\eta}{1-(2+\sqrt{2})\eta}, \quad c_0 = 4+2\sqrt{2}.$$

Proof. Write $a = 1/\theta_0 = 2+\sqrt{2}$ and $b = 1/(\frac{1}{2}-\theta_0) = 2+2\sqrt{2}$, and note $\theta_0 b = \sqrt{2}$, $a\eta_0 = \frac{1}{2}$ and $b\eta_0 = \frac{1}{\sqrt{2}}$. Apply Proposition 4.2 with $G = H = f$, $p = \rho = \frac{1}{2}$ and $\theta = \theta_0 \in (0, \frac{1}{2})$; it gives hypothesis (a) of Lemma 4.3 with $\kappa_a = a\eta$ and hypothesis (b) with $\kappa_b = b\eta$, because $\eta_{f|f} \le \eta$. Hypothesis (c) with $\theta = \theta_0$ is the definition of C. As $\eta \le \eta_0$ we have $\kappa_a \le \frac{1}{2}$ and $\kappa_b \le \frac{1}{\sqrt{2}}$, so Lemma 4.3 applies and bounds $\Pr[f \ne \mathrm{AND}_C]/\mu$ by the largest of

$$T_1 = \frac{a\eta}{1-a\eta}, \qquad T_2 = 2b\eta - 1 + \frac{(1-b\eta)^2 + \theta_0 b\eta}{1-a\eta}, \qquad T_3 = 2b\eta - 1 + \theta_0.$$

Expanding, $(2b\eta-1)(1-a\eta)+(1-b\eta)^2+\theta_0 b\eta = (a+\theta_0 b)\eta + (b^2-2ab)\eta^2 = (2+2\sqrt{2})\eta(1-2\eta)$, since $a+\theta_0 b = 2+2\sqrt{2}$ and $b^2-2ab = b(b-2a) = -2b$. Hence $T_2 = h(\eta)\eta$. Next, $T_2 - T_1 = \eta\frac{\sqrt{2}-(4+4\sqrt{2})\eta}{1-a\eta} \ge 0$ because $(4+4\sqrt{2})\eta_0 = \sqrt{2}$. Also $(1-a\eta)(T_2-T_3) = (12+8\sqrt{2})\eta^2 - (3+3\sqrt{2})\eta + \frac{1}{\sqrt{2}} > 0$, as the discriminant $(3+3\sqrt{2})^2 - 4(12+8\sqrt{2})/\sqrt{2} = -5-6\sqrt{2}$ is negative. So $\Pr[f \ne \mathrm{AND}_C] \le h(\eta)\eta\mu$. Finally, $h'(\eta)$ has the sign of $-2(1-a\eta)+a(1-2\eta) = a-2 > 0$, so $h(\eta) \le h(\eta_0) = (2+2\sqrt{2})\frac{1/\sqrt{2}}{1/2} = 4+2\sqrt{2}$. □

Proof of Theorem 1.1. Put $\eta = \varepsilon/\mu^2$; by Corollary 3.2, $\eta_{f|f} \le \eta$. If $\eta \le \eta_0$, Lemma 5.2 gives $\Pr[f \ne \mathrm{AND}_C] \le h(\eta)\varepsilon/\mu \le c_0\varepsilon/\mu$, which is the second claim. If $\eta > \eta_0 = 1/c_0$ then $c_0\varepsilon/\mu = c_0\eta\mu > \mu = \Pr[f \ne 0]$. In both cases $\mathrm{dist}(f) \le \min\{\mu, c_0\varepsilon/\mu\}$, and $\min\{\mu, a/\mu\} \le \sqrt{a}$ for all $a, \mu > 0$. □

Remark 5.3 (Relative form). For small $\eta = \varepsilon/\mu^2$ Theorem 1.1 reads $\Pr[f \ne \mathrm{AND}_C] \le (2+2\sqrt{2}+O(\eta))\,\varepsilon/\mu$: the *relative* distance $\Pr[f \ne \mathrm{AND}_C]/\mu$ is at most linear in the relative error ε/μ^2. This is sharp up to the constant factor. Let $T, U \subseteq [n]$ be disjoint and nonempty, $a = 2^{-|T|}$, $b = 2^{-|U|}$, and $f = \mathrm{AND}_T \vee \mathrm{AND}_U$. Then $\mu = a+b-ab$, and since f is monotone, $\varepsilon(f) = \mu^2 - \Pr[f(\mathbf{x}\wedge\mathbf{y}) = 1] = \mu^2 - (a^2+b^2-a^2b^2) = 2ab(1-a)(1-b)$. Moreover $\Pr[f \ne \mathrm{AND}_T] = b(1-a)$, while every $g \in \mathcal{A}_n \setminus \{\mathrm{AND}_T\}$ satisfies $\Pr[g \ne \mathrm{AND}_T] \ge a/2$ (for $g = 0$ this is a; for $g = \mathrm{AND}_S$ with $S \supsetneq T$ it is $a(1-2^{-|S\setminus T|}) \ge a/2$; for $S \subsetneq T$ it is $2^{-|S|}(1-2^{-|T\setminus S|}) \ge a$; and for incomparable S, T it is $2^{-|S|}+2^{-|T|}-2^{1-|S\cup T|} \ge \max(2^{-|S|}, 2^{-|T|}) \ge a$ because $|S\cup T| > \max(|S|,|T|)$). So if $b(1-a) \le a/4$, the triangle inequality gives $\mathrm{dist}(f) = b(1-a)$, and

$$\frac{\mathrm{dist}(f)/\mu}{\varepsilon/\mu^2} = \frac{a+b-ab}{2a(1-b)} \;\longrightarrow\; \frac{1}{2} \qquad (|U| \to \infty \text{ with } T \text{ fixed}),$$

while $\varepsilon/\mu^2 \to 0$.

The example in Remark 5.3 shows that the constant in the relative form cannot go below $\frac{1}{2}$. Near the ANDs, $\frac{1}{2}$ is also an upper bound.

Theorem 5.4 (Sharp constant near the ANDs). *Let $f\colon \{0,1\}^n \to \{0,1\}$, $C \subseteq [n]$, $q = 2^{-|C|}$, $\delta_C = \Pr[f \ne \mathrm{AND}_C]$ and $r = \delta_C/q$. Then*

$$\varepsilon(f) \;\ge\; 2q\,\delta_C\,\bigl(1 - 4\sqrt{r}\bigr).$$

Consequently, if $\mu > 0$, $\eta = \varepsilon/\mu^2 < \eta_0$ and C is the set of Theorem 1.1, then $r \le r(\eta) := h(\eta)\eta/(1-h(\eta)\eta)$, and whenever $r(\eta) < \frac{1}{16}$,

$$\operatorname{dist}(f) \;\le\; \Pr[f \neq \mathrm{AND}_C] \;\le\; \frac{1+r(\eta)}{2\,(1-4\sqrt{r(\eta)})}\cdot\frac{\varepsilon}{\mu} \;=\; \Big(\frac{1}{2}+O(\sqrt{\eta})\Big)\frac{\varepsilon}{\mu}.$$

The constant $\frac{1}{2}$ is attained: for $C \cap T = \emptyset$, $|T| \ge 2$ and $f = \mathrm{AND}_C \cdot (1 - \mathrm{AND}_T)$ one has $\operatorname{dist}(f)\,\mu/\varepsilon = \frac{1}{2}$ exactly, while $\varepsilon/\mu^2 = 2^{1-|T|}/(1-2^{-|T|}) \to 0$ as $|T| \to \infty$. Hence

$$\lim_{\eta\to 0}\;\sup\Big\{\frac{\operatorname{dist}(f)\,\mu}{\varepsilon(f)} \;:\; 0<\varepsilon(f)\le \eta\,\mu^2\Big\} \;=\; \frac{1}{2}.$$

The first inequality needs an error term that is at least 1 at $r = \frac{1}{2}$: for $i \in S$, $f = \mathrm{AND}_S$ and $C = S \setminus \{i\}$ give $\varepsilon(f) = 0$ and $r = \frac{1}{2}$. The proof uses one estimate on meets, proved with small-set expansion.

Lemma 5.5 (Meets of small sets)**.** *Let $k \ge 0$, let $\mathbf{x}, \mathbf{y}$ be independent and uniform on $\{0,1\}^k$, $\mathbf{w} = \mathbf{x} \wedge \mathbf{y}$, and let $T \subseteq \{0,1\}^k$ have $\Pr[\mathbf{x} \in T] = d$. Then*

$$4\Pr[\mathbf{x} \in T,\ \mathbf{w} \in T] - \Pr[\mathbf{w} \in T] \;\le\; 4\,d^{3/2}.$$

Proof. Given $\mathbf{w}$, the coordinates of $\mathbf{x}$ are independent, with $\mathbf{x}_i = 1$ if $\mathbf{w}_i = 1$ and $\Pr[\mathbf{x}_i = 1 \mid \mathbf{w}] = \frac{1}{3}$ if $\mathbf{w}_i = 0$ (given $\mathbf{x}_i \wedge \mathbf{y}_i = 0$, the pair $(\mathbf{x}_i, \mathbf{y}_i)$ is uniform on $\{00, 01, 10\}$). Let $U(w) = \Pr[\mathbf{x} \in T \mid \mathbf{w} = w]$. For $t \in \{0,1\}$ and $u \in \mathbb{R}$ we have $t(4u-1) \le 4u^2$, because $4u^2 - 4u + 1 = (2u-1)^2 \ge 0$. Hence

$$4\Pr[\mathbf{x} \in T, \mathbf{w} \in T] - \Pr[\mathbf{w} \in T] = \mathbb{E}\big[\mathbf{1}_T(\mathbf{w})\,(4U(\mathbf{w}) - 1)\big] \le 4\,\mathbb{E}[U(\mathbf{w})^2].$$

Now $\mathbb{E}[U(\mathbf{w})^2] = \Pr[\mathbf{x} \in T,\ \mathbf{x}' \in T]$, where $\mathbf{x}, \mathbf{x}'$ are conditionally independent given $\mathbf{w}$ with the conditional law above. The pairs $(\mathbf{x}_i, \mathbf{x}'_i)$ are independent, with uniform marginals ($\frac{1}{4} + \frac{3}{4}\cdot\frac{1}{3} = \frac{1}{2}$) and $\Pr[\mathbf{x}_i = \mathbf{x}'_i = 1] = \frac{1}{4} + \frac{3}{4}\cdot\frac{1}{9} = \frac{1}{3}$, so $(\mathbf{x}, \mathbf{x}')$ is a $\frac{1}{3}$-correlated pair. Lemma 8.15 gives $\mathbb{E}[U(\mathbf{w})^2] \le d^{2/(1+1/3)} = d^{3/2}$. □

Proof of Theorem 5.4. Write $A = f^{-1}(1)$, $Q = \{x :\ x_C = 1^C\}$, $D = Q \setminus A$, $B = A \setminus Q$, $\delta_D = \Pr[D]$, $\delta_B = \Pr[B]$, so $\delta_C = \delta_D + \delta_B$ and $\mu = q - \delta_D + \delta_B \le q(1+r)$. Every point lies in exactly one of the classes D, $Q \setminus D$, B, $Q^c \setminus B$. The following events are failures ($f(\mathbf{x}\wedge\mathbf{y}) \neq f(\mathbf{x})f(\mathbf{y})$), and they are pairwise disjoint because they prescribe different pairs of classes for $(\mathbf{x}, \mathbf{y})$:

- E_1: $\mathbf{x} \in D$, $\mathbf{y} \in Q \setminus D$, $\mathbf{x}\wedge\mathbf{y} \notin D$ (then $\mathbf{x}\wedge\mathbf{y} \in Q \setminus D$, so $f(\mathbf{x}\wedge\mathbf{y}) = 1 \neq 0 = f(\mathbf{x})$); and E'_1, the same with $\mathbf{x}, \mathbf{y}$ exchanged.
- E_2: $\mathbf{x} \in B$, $\mathbf{y} \in Q \setminus D$, $\mathbf{x}\wedge\mathbf{y} \notin B$ (then $f(\mathbf{x})f(\mathbf{y}) = 1$, and $\mathbf{x}\wedge\mathbf{y} \notin Q$ since $\mathbf{x} \notin Q$, so $f(\mathbf{x}\wedge\mathbf{y}) = 0$); and E'_2, exchanged.
- E_3: $\mathbf{x}, \mathbf{y} \in Q \setminus D$, $\mathbf{x}\wedge\mathbf{y} \in D$.
- E_4: $\mathbf{x} \in Q^c \setminus B$, $\mathbf{y} \in Q$, $\mathbf{x}\wedge\mathbf{y} \in B$ (then $f(\mathbf{x}) = 0$ and $f(\mathbf{x}\wedge\mathbf{y}) = 1$); and E'_4, exchanged.

Let $a_D = \Pr[\mathbf{x} \in D, \mathbf{y} \in Q, \mathbf{x}\wedge\mathbf{y} \in D]$, $b_D = \Pr[\mathbf{x} \in Q, \mathbf{y} \in Q, \mathbf{x}\wedge\mathbf{y} \in D]$, $a_B = \Pr[\mathbf{x} \in B, \mathbf{y} \in Q, \mathbf{x}\wedge\mathbf{y} \in B]$ and $b_B = \Pr[\mathbf{x} \notin Q, \mathbf{y} \in Q, \mathbf{x}\wedge\mathbf{y} \in B]$. By inclusion–exclusion,

$$\Pr[E_1] \ge q\delta_D - \delta_D^2 - a_D, \quad \Pr[E_3] \ge b_D - 2a_D, \quad \Pr[E_2] \ge q\delta_B - \delta_B\delta_D - a_B, \quad \Pr[E_4] = b_B - a_B,$$

and the same for the exchanged events. Summing,

$$\varepsilon \;\geq\; 2q\delta_C - 2\delta_D^2 - 2\delta_B\delta_D + (b_D - 4a_D) + 2(b_B - 2a_B). \tag{3}$$

Let $R = [n] \setminus C$. Conditioned on $\mathbf{x}, \mathbf{y} \in Q$, the restrictions $\mathbf{x}_R, \mathbf{y}_R$ are independent and uniform, and $\mathbf{x} \wedge \mathbf{y} \in Q$. Applying Lemma 5.5 on $\{0,1\}^R$ to $T = \{z : \; (1^C, z) \in D\}$, which has density δ_D/q, gives $b_D - 4a_D = -q^2\big(4\Pr[\mathbf{x}_R \in T, \mathbf{w}_R \in T] - \Pr[\mathbf{w}_R \in T]\big) \geq -4q^2(\delta_D/q)^{3/2}$. For the B terms, split B by the pattern $u = x_C \neq 1^C$: $B_u = \{x \in B : \; x_C = u\}$, $T_u = \{z : \; (u, z) \in B\}$, of density β_u in $\{0,1\}^R$, so $\sum_u \beta_u = \delta_B/q$. If $\mathbf{y} \in Q$ and $\mathbf{x}_C = u$, then $(\mathbf{x} \wedge \mathbf{y})_C = u$. Hence $a_B = q^2 \sum_u a_u$ and $b_B = q^2 \sum_u b_u$, where $a_u = \Pr[\mathbf{x}_R \in T_u, \mathbf{w}_R \in T_u]$ and $b_u = \Pr[\mathbf{w}_R \in T_u]$. Lemma 5.5 gives $b_u - 2a_u \geq 2a_u - 4\beta_u^{3/2} \geq -4\beta_u^{3/2}$, so $b_B - 2a_B \geq -4q^2 \sum_u \beta_u^{3/2} \geq -4q^2(\delta_B/q)^{3/2}$ (superadditivity of $t \mapsto t^{3/2}$). Put $x = \delta_D/q$, so that $\delta_B/q = r - x$ and $2\delta_D^2 + 2\delta_B\delta_D = 2\delta_D\delta_C = 2q^2xr$. Then (3) becomes

$$\varepsilon \;\geq\; q^2\big(2r - L(x)\big), \qquad L(x) = 2xr + 4x^{3/2} + 8(r-x)^{3/2} \quad (0 \leq x \leq r).$$

L is convex, so $L(x) \leq \max\{L(0), L(r)\} = \max\{8r^{3/2},\, 2r^2 + 4r^{3/2}\}$, and this is $8r^{3/2}$ when $r \leq 4$. Hence $\varepsilon \geq 2q^2r - 8q^2r^{3/2} = 2q\delta_C(1 - 4\sqrt{r})$ if $r \leq 4$; if $r > 4$ the right side is negative.

For the consequence, let C be the set of Theorem 1.1. Since h is increasing (proof of Lemma 5.2) and $\eta < \eta_0$, $h(\eta)\eta < h(\eta_0)\eta_0 = 1$, and Theorem 1.1 gives $\delta_C \leq h(\eta)\eta\mu < \mu$. As $q \geq \mu - \delta_B \geq \mu - \delta_C > 0$ and $t \mapsto t/(\mu - t)$ is increasing on $[0, \mu)$, $r \leq h(\eta)\eta/(1 - h(\eta)\eta) = r(\eta) = O(\eta)$. If $\varepsilon = 0$ then $r \leq r(0) = 0$, so $\delta_C = 0$. If $\varepsilon > 0$ and $r(\eta) < \frac{1}{16}$, then $1 - 4\sqrt{r} \geq 1 - 4\sqrt{r(\eta)} > 0$, and the first part with $\mu \leq q(1 + r)$ gives $\delta_C\mu/\varepsilon \leq (1 + r)/(2(1 - 4\sqrt{r})) \leq (1 + r(\eta))/(2(1 - 4\sqrt{r(\eta)}))$, the last step because this function of r is increasing on $[0, \frac{1}{16})$.

For the example, let $f = \mathrm{AND}_C(1 - \mathrm{AND}_T)$ and $\tau = 2^{-|T|}$. Then $B = \emptyset$ and $D = Q \cap \{x_T = 1^T\}$, and $f = 1$ exactly on $G := Q \setminus D = Q \cap A$. The set D is closed under meets, and it is an up-set: if $z \in D$ and $x \geq z$ then $x \in D$. The failures are exactly the pairs in $E_1 \cup E_1'$. Indeed, if $\mathbf{x} \notin Q$ or $\mathbf{y} \notin Q$ then $\mathbf{x} \wedge \mathbf{y} \notin Q$, so $f(\mathbf{x} \wedge \mathbf{y}) = 0 = f(\mathbf{x})f(\mathbf{y})$. Let $\mathbf{x}, \mathbf{y} \in Q$, so that $\mathbf{x} \wedge \mathbf{y} \in Q$. If $\mathbf{x}, \mathbf{y} \in D$ then $\mathbf{x} \wedge \mathbf{y} \in D$, and $f(\mathbf{x} \wedge \mathbf{y}) = 0 = f(\mathbf{x})f(\mathbf{y})$. If $\mathbf{x}, \mathbf{y} \in G$ then $\mathbf{x} \wedge \mathbf{y} \notin D$, since $\mathbf{x} \geq \mathbf{x} \wedge \mathbf{y}$, $\mathbf{x} \notin D$ and D is an up-set; so $\mathbf{x} \wedge \mathbf{y} \in G$ and $f(\mathbf{x} \wedge \mathbf{y}) = 1 = f(\mathbf{x})f(\mathbf{y})$. If $\mathbf{x} \in D$ and $\mathbf{y} \in G$ then $f(\mathbf{x})f(\mathbf{y}) = 0$, while $(\mathbf{x} \wedge \mathbf{y})_T = \mathbf{y}_T \neq 1^T$ gives $\mathbf{x} \wedge \mathbf{y} \in G$ and $f(\mathbf{x} \wedge \mathbf{y}) = 1$: all these pairs fail, and they form E_1 (and the exchanged pairs form E_1'). Hence $\varepsilon = 2\Pr[D]\Pr[G] = 2q^2\tau(1 - \tau)$, while $\mu = q(1 - \tau)$ and $\Pr[f \neq \mathrm{AND}_C] = q\tau$. Every $g \in \mathcal{A}_n \setminus \{\mathrm{AND}_C\}$ has $\Pr[g \neq \mathrm{AND}_C] \geq q/2$ (Remark 5.3). So for $\tau \leq \frac{1}{4}$ the triangle inequality gives $\mathrm{dist}(f) = q\tau$ and $\mathrm{dist}(f)\mu/\varepsilon = \frac{1}{2}$. Combined with the first part, this gives the limit. □

The same argument gives the sharp relative constant for the m-ary test of Theorem 1.4(b).

Theorem 5.6 (Sharp constant for the m-ary test)**.** *Let $m \geq 2$, $\varepsilon_m = \Pr[f(\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_m) \neq \prod_j f(\mathbf{x}_j)]$ for independent uniform $\mathbf{x}_j$, and let C, q, δ_C, r be as in Theorem 5.4. With $\kappa_m = 2 - 2^{1-m}$,*

$$\varepsilon_m \;\geq\; m\, q^{m-1}\delta_C\,\big(1 - m^2\, r^{\kappa_m - 1}\big).$$

Consequently, if $\mu > 0$ and $\eta_m = \varepsilon_m/\mu^m$ satisfies $c_0\eta_m < 1$, then $\mathrm{dist}(f) = \Pr[f \neq \mathrm{AND}_C]$ for some C with $r \leq r_m := c_0\eta_m/(1 - c_0\eta_m)$, and

$$\mathrm{dist}(f) \;\leq\; \frac{(1 + r_m)^{m-1}}{m\,\big(1 - m^2 r_m^{\kappa_m - 1}\big)} \cdot \frac{\varepsilon_m}{\mu^{m-1}} \;=\; \Big(\frac{1}{m} + O_m\big(\sqrt{\eta_m}\big)\Big)\frac{\varepsilon_m}{\mu^{m-1}}$$

whenever $m^2 r_m^{\kappa_m - 1} < 1$. The constant $\frac{1}{m}$ is optimal: $f = \mathrm{AND}_C(1 - \mathrm{AND}_T)$ with $C \cap T = \emptyset$ and $\tau = 2^{-|T|} \leq \frac{1}{4}$ has $\mathrm{dist}(f)\,\mu^{m-1}/\varepsilon_m = \tau(1 - \tau)^{m-1}/(1 - (1 - \tau)^m - \tau^m) \to \frac{1}{m}$ as $|T| \to \infty$.

Proof. Meet lemma. Let $\mathbf{x}$ be uniform and $\mathbf{y}'$ be $\mu_{2^{1-m}}$-distributed on $\{0,1\}^k$, independent, $\mathbf{w} = \mathbf{x} \wedge \mathbf{y}'$, and $\Pr[\mathbf{x} \in T] = d$. Then

$$2m \Pr[\mathbf{x} \in T, \mathbf{w} \in T] - \Pr[\mathbf{w} \in T] \le m^2 d^{\kappa_m}.$$

The proof is that of Lemma 5.5, with these changes. Given $\mathbf{w}$, the coordinates of $\mathbf{x}$ are independent, with $\mathbf{x}_i = 1$ if $\mathbf{w}_i = 1$ and $\Pr[\mathbf{x}_i = 1 \mid \mathbf{w}_i = 0] = (\frac{1}{2} - s)/(1-s)$, where $s = 2^{-m}$. The pointwise inequality becomes $t(2mu - 1) \le m^2 u^2$, since $(mu-1)^2 \ge 0$. Two conditionally independent copies $\mathbf{x}, \mathbf{x}'$ have uniform marginals and $\Pr[\mathbf{x}_i = \mathbf{x}'_i = 1] = s + (\frac{1}{2} - s)^2/(1-s) = \frac{1}{4(1-s)}$. So $(\mathbf{x}, \mathbf{x}')$ is ρ_m-correlated with $\rho_m = s/(1-s) = 1/(2^m - 1)$, and Lemma 8.15 gives $\mathbb{E}[U(\mathbf{w})^2] \le d^{2/(1+\rho_m)} = d^{2-2^{1-m}} = d^{\kappa_m}$.

Accounting. With the four classes D, $G = Q \setminus D$, B, $N = Q^c \setminus B$ as before, the following events are failures and pairwise disjoint (different class patterns of $(\mathbf{x}_1, \dots, \mathbf{x}_m)$), where $\mathbf{w} = \mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_m$:

- $F_1(j)$: $\mathbf{x}_j \in D$, $\mathbf{x}_i \in G$ for $i \ne j$, and $\mathbf{w} \notin D$;
- $F_2(j)$: $\mathbf{x}_j \in B$, $\mathbf{x}_i \in G$ for $i \ne j$, and $\mathbf{w} \notin B$;
- F_3: all $\mathbf{x}_i \in G$ and $\mathbf{w} \in D$;
- $F_4(j)$: $\mathbf{x}_j \in N$, $\mathbf{x}_i \in Q$ for $i \ne j$, and $\mathbf{w} \in B$.

Let $a_D = \Pr[\mathbf{x}_1 \in D,\ \mathbf{x}_2, \dots, \mathbf{x}_m \in Q,\ \mathbf{w} \in D]$, $b_D = \Pr[\mathbf{x}_1, \dots, \mathbf{x}_m \in Q,\ \mathbf{w} \in D]$, $a_B = \Pr[\mathbf{x}_1 \in B,\ \mathbf{x}_2, \dots, \mathbf{x}_m \in Q,\ \mathbf{w} \in B]$ and $b_B = \Pr[\mathbf{x}_1 \notin Q,\ \mathbf{x}_2, \dots, \mathbf{x}_m \in Q,\ \mathbf{w} \in B]$. Then

$$\Pr[F_1(j)] \ge q^{m-1}\delta_D - (m-1)q^{m-2}\delta_D^2 - a_D, \qquad \Pr[F_3] \ge b_D - m\,a_D,$$
$$\Pr[F_2(j)] \ge q^{m-1}\delta_B - (m-1)q^{m-2}\delta_B\delta_D - a_B, \qquad \Pr[F_4(j)] = b_B - a_B,$$

so $\varepsilon_m \ge mq^{m-1}\delta_C - m(m-1)q^{m-2}\delta_D\delta_C + (b_D - 2m\,a_D) + m(b_B - 2a_B)$. Given $\mathbf{x}_2, \dots, \mathbf{x}_m \in Q$, the restriction to $R = [n] \setminus C$ of $\mathbf{x}_2 \wedge \cdots \wedge \mathbf{x}_m$ is $\mu_{2^{1-m}}$-distributed. So the meet lemma, applied as in the proof of Theorem 5.4, gives $b_D - 2ma_D \ge -m^2 q^m (\delta_D/q)^{\kappa_m}$. Per pattern u it gives $b_u - 2a_u \ge (2m-2)a_u - m^2\beta_u^{\kappa_m}$, hence $b_B - 2a_B \ge -m^2 q^m (\delta_B/q)^{\kappa_m}$ by superadditivity of $t \mapsto t^{\kappa_m}$. With $x = \delta_D/q \in [0, r]$ this gives $\varepsilon_m \ge q^m(mr - L(x))$, where $L(x) = m(m-1)xr + m^2 x^{\kappa_m} + m^3(r-x)^{\kappa_m}$ is convex. If $r \le 1$, then $L(r) = m(m-1)r^2 + m^2 r^{\kappa_m} \le m^3 r^{\kappa_m} = L(0)$, because $r^{2-\kappa_m} \le 1 \le m$; so $L \le m^3 r^{\kappa_m}$ and $\varepsilon_m \ge mq^{m-1}\delta_C(1 - m^2 r^{\kappa_m - 1})$. If $r > 1$ the right side is negative.

Consequence. If $\varepsilon_m = 0$, Theorem 1.4(b) gives $\mathrm{dist}(f) = 0$. Otherwise, by Theorem 1.4(b), $\mathrm{dist}(f) \le c_0 \eta_m \mu < \mu = \Pr[f \ne 0]$. So $\mathrm{dist}(f) = \delta_C$ for some C, and $q \ge \mu - \delta_C \ge \mu(1 - c_0\eta_m)$, which gives $r \le r_m$. Hence $1 - m^2 r^{\kappa_m - 1} \ge 1 - m^2 r_m^{\kappa_m - 1} > 0$. Also $\mu^{m-1} \le q^{m-1}(1+r)^{m-1}$, so $\mathrm{dist}(f)\mu^{m-1}/\varepsilon_m$ is at most the displayed factor at r, which is increasing in r. Finally $\kappa_m - 1 = 1 - 2^{1-m} \ge \frac{1}{2}$ and $r_m \le 1$, so $r_m^{\kappa_m - 1} \le \sqrt{r_m} = O(\sqrt{\eta_m})$.

Example. Here $B = \emptyset$, $D = Q \cap \{x_T = 1^T\}$, and $f = 1$ exactly on $G = Q \setminus D = Q \cap A$; as in the proof of Theorem 5.4, D is an up-set. If some $\mathbf{x}_j \notin Q$, then $\mathbf{w} \notin Q$ and $f(\mathbf{w}) = 0 = \prod_j f(\mathbf{x}_j)$. If all $\mathbf{x}_j \in G$, then $\mathbf{w} \in Q$, and $\mathbf{w} \notin D$ because $\mathbf{x}_1 \ge \mathbf{w}$, $\mathbf{x}_1 \notin D$ and D is an up-set; so $f(\mathbf{w}) = 1 = \prod_j f(\mathbf{x}_j)$. If all $\mathbf{x}_j \in Q$ and some $\mathbf{x}_j \in D$, then $\prod_j f(\mathbf{x}_j) = 0$, and since $\mathbf{w}_T = 1^T$ iff $(\mathbf{x}_j)_T = 1^T$ for every j, the tuple fails iff some $\mathbf{x}_j$ lies in G. So a tuple fails iff all $\mathbf{x}_j \in Q$ and both D and G contain some $\mathbf{x}_j$. Given that all $\mathbf{x}_j \in Q$, the events $\{\mathbf{x}_j \in D\}$ are independent, each of probability τ, so $\varepsilon_m = q^m(1 - (1-\tau)^m - \tau^m)$, while $\mu = q(1-\tau)$ and $\mathrm{dist}(f) = q\tau$ (as in the proof of Theorem 5.4). □

The script `sharp_relative.py` (directory `numerics`) checks Lemma 5.5 and the meet lemma for $m = 3$ on random and structured sets, (3) and the first inequalities of Theorems 5.4 and 5.6 ($m = 3$) for every C on small n, both first inequalities in exact rational arithmetic for perturbations $\mathrm{AND}_C \oplus \mathbf{1}_E$ with r in the non-vacuous range and $n \le 20$ (with ε and ε_3 computed exactly by Möbius inversion), the consequence of Theorem 5.4, and the example.

Proof of Theorem 1.2. Slices. Let $f_n = \mathbf{1}\{|x| = n-1\}$. Then $f_n(x)f_n(y) = 1$ iff $|x| = |y| = n-1$, and $f_n(x \wedge y) = 1$ iff $(x, y) \in \{(s, s), (s, 1^n), (1^n, s)\}$ for some s with $|s| = n-1$. The failing pairs are therefore the $n(n-1)$ pairs of distinct points of the slice and the $2n$ pairs $(s, 1^n), (1^n, s)$, so $\varepsilon(f_n) = n(n+1)/4^n$. For $S \subseteq [n]$ with $u = n - |S|$, $\Pr[f_n = 1, \mathrm{AND}_S = 1] = u/2^n$ (the single zero must lie outside S), so $\Pr[f_n \ne \mathrm{AND}_S] = (n + 2^u - 2u)/2^n \ge n/2^n = \Pr[f_n \ne 0]$ because $2^u \ge 2u$ for all integers $u \ge 0$. Hence $\mathrm{dist}(f_n) = n/2^n$ and $\mathrm{dist}(f_n)^2/\varepsilon(f_n) = n/(n+1)$.

Monotone functions. If f is monotone then $f(x \wedge y) \le f(x)f(y)$, so $\varepsilon(f) = \mathbb{E}[f(\mathbf{x})f(\mathbf{y}) - f(\mathbf{x} \wedge \mathbf{y})] \le \mu^2$.

Thresholds. Fix $0 < p < 1/2$ and $n \ge 1$. Let t be the least integer with $\Pr[|\mathbf{x}| \ge t] \le p$; then $1 \le t \le n+1$ and $\Pr[|\mathbf{x}| \ge t-1] > p$. Choose a set L of points with $|x| = t-1$ such that $U = \{x : |x| \ge t\} \cup L$ satisfies $p - 2^{-n} < \mu(U) \le p$, and let $f = \mathbf{1}_U$. U is an up-set (every point strictly above a point of level $t-1$ has weight $\ge t$), so $\varepsilon(f) \le \mu(U)^2 \le p^2$.

We bound $\mathrm{dist}(f)$ from below. $\Pr[f \ne 0] = \mu(U) > p - 2^{-n}$ and $\Pr[f \ne \mathrm{AND}_\emptyset] = 1 - \mu(U) > 1/2 > p$. Let $S \ne \emptyset$, $s = |S|$, and $q_S = \Pr[f = 1 \mid \mathbf{x}_S = 1^S]$. Then

$$\Pr[f \ne \mathrm{AND}_S] = \mu(U) + 2^{-s} - 2 \cdot 2^{-s} q_S = \mu(U) - 2^{-s}(2q_S - 1).$$

If $q_S \le 1/2$ this is $\ge \mu(U)$. If $q_S > 1/2$, then, as $f = 1$ implies $|\mathbf{x}| \ge t-1$, we get $\Pr[\mathrm{Bin}(n-s, \frac{1}{2}) \ge t-1-s] \ge q_S > 1/2$; by the symmetry of $\mathrm{Bin}(N, \frac{1}{2})$ about $N/2$, $\Pr[\mathrm{Bin}(N, \frac{1}{2}) \ge a] \le 1/2$ whenever $a > N/2$, so $t-1-s \le (n-s)/2$, i.e. $s \ge 2t-2-n$, and then $\Pr[f \ne \mathrm{AND}_S] \ge \mu(U) - 2^{-(2t-2-n)}$.

It remains to show $2t - n \to \infty$ as $n \to \infty$. Every point mass of $\mathrm{Bin}(n, \frac{1}{2})$ is at most $\beta_n := \binom{n}{\lfloor n/2 \rfloor} 2^{-n} \le 1/\sqrt{n}$. (Let $a_k = \binom{2k}{k} 4^{-k} = \prod_{j=1}^{k} \frac{2j-1}{2j}$. Since $\frac{2j-1}{2j} \le \frac{2j}{2j+1}$, we get $a_k^2 \le \prod_{j=1}^{k} \frac{2j-1}{2j} \cdot \frac{2j}{2j+1} = \frac{1}{2k+1}$. Hence $\beta_{2k} = a_k \le 1/\sqrt{2k}$ for $k \ge 1$, and $\beta_{2k+1} = a_k \cdot \frac{2k+1}{2k+2} \le a_k \le 1/\sqrt{2k+1}$ for $k \ge 0$, using $\binom{2k+1}{k} = \binom{2k}{k} \frac{2k+1}{k+1}$.) Let $k = \lfloor (\frac{1}{2} - p)\sqrt{n}/2 \rfloor$. Since $\Pr[|\mathbf{x}| \ge \lceil n/2 \rceil] \ge \frac{1}{2}$ (the events $|\mathbf{x}| \ge \lceil n/2 \rceil$ and $|\mathbf{x}| \le \lfloor n/2 \rfloor$ have equal probability and cover the cube),

$$\Pr\left[|\mathbf{x}| \ge \lceil n/2 \rceil + k\right] \;\ge\; \frac{1}{2} - \frac{k}{\sqrt{n}} \;\ge\; \frac{1}{2} - \frac{1/2 - p}{2} \;>\; p,$$

so $t > \lceil n/2 \rceil + k$ and $2t - 2 - n \ge 2k \to \infty$. Altogether

$$\varepsilon(f) \le p^2 \qquad \text{and} \qquad \mathrm{dist}(f) \;\ge\; p - 2^{-n} - 2^{-2k} \xrightarrow[n \to \infty]{} p.$$

Hence $\delta^*(p^2) \ge p$ for every $p \in (0, 1/2)$, i.e. $\delta^*(\varepsilon) \ge \sqrt{\varepsilon}$ for $0 < \varepsilon < 1/4$; for $\varepsilon = 1/4$ use monotonicity of δ^*. □

Proof of Corollary 1.3. Combine Theorems 1.1 and 1.2. □

6 Extensions

Proof of Theorem 1.4(a). This is Theorem 5.1. □

Proof of Theorem 1.4(b). By Corollary 3.3 (with $p = 1/2$), $\eta_{f|f} \le \eta := \varepsilon_m/\mu^m$. If $\eta \le \eta_0$, Lemma 5.2 gives $\Pr[f \ne \mathrm{AND}_C] \le c_0\eta\mu = c_0\varepsilon_m/\mu^{m-1}$; if $\eta > \eta_0 = 1/c_0$ then $c_0\varepsilon_m/\mu^{m-1} = c_0\eta\mu > \mu = \Pr[f \ne 0]$. Thus $\mathrm{dist}(f) \le \min\{\mu, c_0\varepsilon_m/\mu^{m-1}\}$. If $\mu^m \le c_0\varepsilon_m$ the minimum is at most $\mu \le (c_0\varepsilon_m)^{1/m}$; otherwise it is at most $c_0\varepsilon_m/\mu^{m-1} < (c_0\varepsilon_m)/(c_0\varepsilon_m)^{(m-1)/m} = (c_0\varepsilon_m)^{1/m}$.

For the lower bound, if f is monotone then $f(\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_m) \le \prod_j f(\mathbf{x}_j)$, hence $\varepsilon_m(f) = \mathbb{E}\prod_j f(\mathbf{x}_j) - \mathbb{E}f(\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_m) \le \mu^m$. For $0 < p < 1/2$, the up-sets U from the proof of Theorem 1.2 with density $\mu(U) \in (p - 2^{-n}, p]$ have $\varepsilon_m(\mathbf{1}_U) \le p^m$ and $\mathrm{dist}(\mathbf{1}_U) \ge p - 2^{-n} - 2^{-2k} \to p$ as $n \to \infty$. Taking $p = \varepsilon^{1/m}$ shows that $\delta^*_m(\varepsilon) := \sup_n \sup\{\mathrm{dist}(f) : \varepsilon_m(f) \le \varepsilon\} \ge \varepsilon^{1/m}$ for all $0 < \varepsilon < 2^{-m}$. □

Proof of Theorem 1.4(c). Write $\mu_G = \mathbb{E}_{\mu_p} G$, $\mu_H = \mathbb{E}_{\mu_\rho} H$, and $\eta_0 = (1-p)(1-\rho)/25$, so that $K = 1/\sqrt{\eta_0}$.

Case (i): $\min(\mu_G, \mu_H) \le K\sqrt{\varepsilon}$. Since $\mathbf{x} \wedge \mathbf{y} \sim \mu_{p\rho}$,

$$\begin{aligned}\Pr_{\mu_{p\rho}}[F = 1] = \Pr[F(\mathbf{x}\wedge\mathbf{y}) = 1] &\le \Pr[F(\mathbf{x}\wedge\mathbf{y}) \ne G(\mathbf{x})H(\mathbf{y})] + \Pr[G(\mathbf{x})H(\mathbf{y}) = 1] \\ &\le \varepsilon + \min(\mu_G, \mu_H) \le \varepsilon + K\sqrt{\varepsilon}.\end{aligned}$$

Case (ii): $\mu_G > K\sqrt{\varepsilon}$ *and* $\mu_H > K\sqrt{\varepsilon}$. In particular $\mu_G, \mu_H > 0$, and $\mu_G\mu_H \ge K^2\varepsilon = \varepsilon/\eta_0$, so $\eta := \varepsilon/(\mu_G\mu_H) \le \eta_0$. By Lemma 3.1 and symmetry, $\eta_{G|H} \le \eta$ and $\eta_{H|G} \le \eta$.

Apply Proposition 4.2 to (G, H) with $\theta = (1-\rho)/2$, obtaining $C_1 = \{i : P_H[\mathbf{y}_i = 0] \le (1-\rho)/2\}$ and $\kappa_1 = 2\eta/(1-\rho) \le 2\eta_0/(1-\rho) \le 2/25 < 1/3$:

$$P_G[\mathbf{x}_{C_1} \ne 1^{C_1}] \le \kappa_1, \qquad \Pr_{\mu_p}[G \ne \mathrm{AND}_{C_1}] \le 3\kappa_1\mu_G.$$

Apply it to (H, G) (roles of p, ρ exchanged) with $\theta = (1-p)/2$, obtaining $C_2 = \{i : P_G[\mathbf{x}_i = 0] \le (1-p)/2\}$, $\kappa_2 = 2\eta/(1-p) < 1/3$ and $\Pr_{\mu_\rho}[H \ne \mathrm{AND}_{C_2}] \le 3\kappa_2\mu_H$.

Claim: $C_1 = C_2$. If $i \in C_1$ then $P_G[\mathbf{x}_i = 0] \le P_G[\mathbf{x}_{C_1} \ne 1^{C_1}] \le \kappa_1 = 2\eta/(1-\rho) \le 2(1-p)/25 < (1-p)/2$, so $i \in C_2$. If $i \notin C_1$, let $r = 3\kappa_1 = 6\eta/(1-\rho) \le 6(1-p)/25$. Since $\Pr[\mathrm{AND}_{C_1} = 1] = p^{|C_1|}$ and $\Pr[G \ne \mathrm{AND}_{C_1}] \le r\mu_G$, we have $p^{|C_1|} \ge (1-r)\mu_G$ and

$$\Pr_{\mu_p}[G = 1, \mathbf{x}_i = 0] \;\ge\; \Pr[\mathrm{AND}_{C_1} = 1, \mathbf{x}_i = 0] - r\mu_G = (1-p)p^{|C_1|} - r\mu_G \;\ge\; \big((1-p)(1-r) - r\big)\mu_G,$$

so $P_G[\mathbf{x}_i = 0] \ge (1-p) - r(2-p) \ge (1-p) - \frac{12}{25}(1-p) > (1-p)/2$, i.e. $i \notin C_2$. This proves the claim; let $T = C_1 = C_2$.

Now, using $\mu_H > K\sqrt{\varepsilon}$ (and reading the middle term as 0 if $\varepsilon = 0$), $\Pr_{\mu_p}[G \ne \mathrm{AND}_T] \le 3\kappa_1\mu_G = \frac{6\varepsilon}{(1-\rho)\mu_H} \le \frac{6\sqrt{\varepsilon}}{(1-\rho)K} = \frac{6}{5}\sqrt{\frac{1-p}{1-\rho}}\sqrt{\varepsilon}$, and symmetrically $\Pr_{\mu_\rho}[H \ne \mathrm{AND}_T] \le \frac{6\varepsilon}{(1-p)\mu_G} \le \frac{6}{5}\sqrt{\frac{1-\rho}{1-p}}\sqrt{\varepsilon}$. Finally, since $\mathrm{AND}_T(\mathbf{x}\wedge\mathbf{y}) = \mathrm{AND}_T(\mathbf{x})\mathrm{AND}_T(\mathbf{y})$ and $\mathbf{x}\wedge\mathbf{y} \sim \mu_{p\rho}$,

$$\Pr_{\mu_{p\rho}}[F \ne \mathrm{AND}_T] \le \Pr[F(\mathbf{x}\wedge\mathbf{y}) \ne G(\mathbf{x})H(\mathbf{y})] + \Pr[G(\mathbf{x}) \ne \mathrm{AND}_T(\mathbf{x})] + \Pr[H(\mathbf{y}) \ne \mathrm{AND}_T(\mathbf{y})],$$

which is at most $\varepsilon + \frac{6}{5}\sqrt{\varepsilon}\big(\sqrt{\frac{1-p}{1-\rho}} + \sqrt{\frac{1-\rho}{1-p}}\big) = \varepsilon + \frac{6}{5}\cdot\frac{2-p-\rho}{\sqrt{(1-p)(1-\rho)}}\sqrt{\varepsilon} \le \varepsilon + \frac{12}{5}\frac{\sqrt{\varepsilon}}{\sqrt{(1-p)(1-\rho)}}$. □

Remark 6.1 (Dictatorship testing). FLMM observe that their theorem shows that the natural dictatorship tester of Parnas, Ron and Samorodnitsky [28] (check $\mathbb{E}f \approx \frac{1}{2}$ and $f(\mathbf{x}\wedge\mathbf{y}) = f(\mathbf{x}) \wedge f(\mathbf{y})$) is sound. Theorem 1.1 makes this quantitative with square-root loss: if $\varepsilon(f) = \varepsilon$ and $|\mathbb{E}f - \frac{1}{2}| + \sqrt{c_0\varepsilon} < \frac{1}{4}$, then f is $\sqrt{c_0\varepsilon}$-close to a dictator x_i. Indeed, the closest $g \in \mathcal{A}_n$ satisfies $|\mathbb{E}g - \mathbb{E}f| \le \sqrt{c_0\varepsilon}$, so $\mathbb{E}g \in (\frac{1}{4}, \frac{3}{4})$, which forces $g = \mathrm{AND}_{\{i\}} = x_i$.

7 On the constant

Theorems 1.1 and 1.2 give $1 \le \delta^*(\varepsilon)/\sqrt{\varepsilon} \le \sqrt{c_0}$. The slices and thresholds of Theorem 1.2 have $\mathrm{dist}^2/\varepsilon \to 1$ from below. Exact enumeration of all f on $n \le 4$ variables gives $\max \mathrm{dist}(f)^2/\varepsilon(f) = n/(n+1)$, attained by the slice, and the maximum over all 2^{n+1} symmetric functions stays below 1 for every $n \le 16$ (0.9804 at $n = 16$). This suggests $\delta^*(\varepsilon) = \sqrt{\varepsilon}$, but that is false.

Lemma 7.1 (Padding)**.** *Let $g\colon \{0,1\}^m \to \{0,1\}$, $k \ge 0$, and $f(z,x) = \mathrm{AND}_{[k]}(z)\, g(x)$ for $(z,x) \in \{0,1\}^k \times \{0,1\}^m$. Then $\varepsilon(f) = 4^{-k}\varepsilon(g)$ and $\mathrm{dist}(f) = 2^{-k}\mathrm{dist}(g)$.*

Proof. Write $a(z) = \mathrm{AND}_{[k]}(z)$ and note $a(z \wedge z') = a(z)a(z')$. Hence $f((z,x) \wedge (z',y)) = a(z)a(z')g(x \wedge y)$ and $f(z,x)f(z',y) = a(z)a(z')g(x)g(y)$. Both vanish unless $a(z) = a(z') = 1$, an event of probability 4^{-k} independent of (x,y); so $\varepsilon(f) = 4^{-k}\varepsilon(g)$. If $h \in \mathcal{A}_m$ is closest to g, then $a \cdot h \in \mathcal{A}_{k+m}$ (it is 0 if $h = 0$ and $\mathrm{AND}_{[k]\cup S}$ if $h = \mathrm{AND}_S$), and f and $a \cdot h$ agree when $z \ne 1^k$, so $\mathrm{dist}(f) \le \Pr[f \ne a \cdot h] = 2^{-k}\Pr[g \ne h] = 2^{-k}\mathrm{dist}(g)$. Conversely, for any $h' \in \mathcal{A}_{k+m}$ the restriction $h'(1^k, \cdot)$ lies in $\mathcal{A}_m$ (a restriction of 0 is 0, and a restriction of an AND is an AND), so $\Pr[f \ne h'] \ge 2^{-k}\Pr_x[g(x) \ne h'(1^k,x)] \ge 2^{-k}\mathrm{dist}(g)$. □

Proposition 7.2 (The constant 1 is not sharp)**.** *Let $n = 69$, $Z = \{0,3\} \cup \{21,22,\dots,33\} \cup \{43,44,\dots,47\}$, and let $f(x) = 1$ iff the number of zeros of x lies in Z. Then*

$$2^n \mathrm{dist}(f) = 245699879978975341077,$$
$$4^n \varepsilon(f) = 537992987945423102030228278419775298689 34,$$

so $\mathrm{dist}(f)^2/\varepsilon(f) = 1.12210\ldots > 1$. Consequently $\delta^*(\varepsilon) \ge \sqrt{1.1221\,\varepsilon} > 1.059\sqrt{\varepsilon}$ for $\varepsilon = 4^{-k}\varepsilon(f)$ and every $k \ge 0$; in particular $\limsup_{\varepsilon\to 0} \delta^*(\varepsilon)/\sqrt{\varepsilon} > 1.059$.

Proof. Let $F(w) \in \{0,1\}$ be the value of f on inputs of Hamming weight w, i.e. $F(w) = 1$ iff $n - w \in Z$. Grouping the pairs (x,y) by the numbers a,b,c,d of coordinates with $(x_i,y_i) = (1,1),(1,0),(0,1),(0,0)$, so that $|x \wedge y| = a$, $|x| = a+b$ and $|y| = a+c$, gives

$$4^n \varepsilon(f) = \sum_{a+b+c+d=n} \binom{n}{a,b,c,d} \mathbf{1}\{F(a) \ne F(a+b)F(a+c)\}.$$

Since f is symmetric, $\Pr[f \ne \mathrm{AND}_S]$ depends only on $s = |S|$; the number of x of weight w with $x_S = 1^S$ is $\binom{n-s}{w-s}$, so with $M = \sum_w \binom{n}{w} F(w) = 2^n \Pr[f \ne 0]$,

$$2^n \mathrm{dist}(f) = \min\Big\{M,\ \min_{0\le s\le n}\Big(M + \sum_{w=s}^{n} \binom{n-s}{w-s}(1 - 2F(w))\Big)\Big\}.$$

Evaluating both expressions in exact integer arithmetic gives the two integers displayed (the minimum is attained at $s = 3$); the script `numerics/counterexample_sharp.py` does this. An independent floating-point evaluation through the closed forms $4^n\varepsilon = \sum_{m\in Z}\binom{n}{m}3^m + M^2 - 2\sum_{m\in Z}\binom{n}{m}N_m$, with $N_m = \#\{(P,Q) : P \cup Q = [m],\ |P|,|Q| \in Z\}$ (here P, Q are the zero sets of x, y), agrees to 10^{-15}. The consequence follows from Lemma 7.1: $\mathrm{AND}_{[k]} \cdot f$ has $\varepsilon = 4^{-k}\varepsilon(f)$ and $\mathrm{dist}^2 = 4^{-k}\mathrm{dist}(f)^2 > 1.1221 \cdot 4^{-k}\varepsilon(f)$. □

The example has two bands of weights. One is around $0.6n$ and carries almost all of the mass ($\mu \approx 0.43$). The other is around $0.35n$, where the meets of two points of the first band typically land. So $f(\mathbf{x} \wedge \mathbf{y}) = f(\mathbf{x})f(\mathbf{y}) = 1$ happens often, while $\mathrm{dist}(f)$ stays close to μ. Larger n gives larger ratios.

Proposition 7.3 (A larger certified ratio)**.** *Let $n = 16000$, $Z = Z_1 \cup Z_2$ with $Z_1 = \{7519, \dots, 7999\}$ and $Z_2 = \{11595, \dots, 11949\}$, and let $f(x) = 1$ iff the number of zeros of x lies in Z. Then $\operatorname{dist}(f)^2/\varepsilon(f) > 1.342$ (the value is $1.342205\dots$, and the member of $\mathcal{A}_n$ closest to f is AND_S with $|S| = 3$). Consequently $\delta^*(\varepsilon) > \sqrt{1.342\,\varepsilon}$ for $\varepsilon = 4^{-k}\varepsilon(f)$, $k \ge 0$, and $\limsup_{\varepsilon\to 0} \delta^*(\varepsilon)/\sqrt{\varepsilon} > 1.1584$.*

Proof. Let $f = \mathbf{1}\{\#\text{zeros} \in Z\}$ for an arbitrary $Z \subseteq \{0, \dots, n\}$ and $M = 2^n\mu = \sum_{m\in Z} \binom{n}{m}$. Since $\mathbf{1}\{a \ne b\} = a + b - 2ab$ for $a, b \in \{0, 1\}$,

$$\varepsilon(f) = \Pr[f(\mathbf{x} \wedge \mathbf{y}) = 1] + \mu^2 - 2\tau, \qquad \tau = \Pr[f(\mathbf{x} \wedge \mathbf{y}) = f(\mathbf{x}) = f(\mathbf{y}) = 1].$$

Each coordinate is a zero of $\mathbf{x}\wedge\mathbf{y}$ with probability $3/4$, so $4^n \Pr[f(\mathbf{x}\wedge\mathbf{y}) = 1] = H := \sum_{m\in Z} \binom{n}{m} 3^m$. The zero set of $\mathbf{x} \wedge \mathbf{y}$ is $P \cup Q$, where P, Q are the zero sets of $\mathbf{x}, \mathbf{y}$. Group the pairs (P, Q) by $U = P \cup Q$, $|U| = m$: choose $P \subseteq U$ with $|P| = i$; then $Q \supseteq U \setminus P$ and $Q \cap P$ is any k-subset of P, so $|Q| = m - i + k$. Hence $4^n\tau = G := \sum_{m\in Z} \binom{n}{m} N_m$ with

$$N_m = \sum_{i\in Z,\ i\le m} \binom{m}{i} \sum_{k=0}^{i} \binom{i}{k} \mathbf{1}\{m - i + k \in Z\}, \qquad \text{and} \qquad 4^n\varepsilon(f) = E := H + M^2 - 2G.$$

For $|S| = n-u$, the points with $f = 1$ and $x_S = 1^S$ have their zeros among the other u coordinates, so, as in the proof of Proposition 7.2,

$$2^n \operatorname{dist}(f) = D := \min\Big\{M,\ \min_{0\le u\le n} \Big(M + 2^u - 2 \sum_{j\in Z,\ j\le u} \binom{u}{j}\Big)\Big\}.$$

The script `numerics/exact_band_check.py` evaluates M, D and E in exact integer arithmetic (rows of Pascal's triangle and prefix sums over the two intervals of Z). For the n and Z of the statement it returns integers with 4817, 4817 and 9633 decimal digits satisfying $1000D^2 > 1342E$ (they are stored in the file `exact_16000_integers.txt` of the folder `numerics`); this takes a few minutes. The same script reproduces the two integers of Proposition 7.2, and for $n \in \{40, 69, 100, 300\}$ it agrees exactly with the quadruple sum used there. The consequences follow from Lemma 7.1 as before, and $\sqrt{1.342} > 1.1584$. □

Proposition 7.4 (The Gaussian limit of two bands)**.** *Let $u_0 < u_1$ be real, and let X, Y, V be independent $N(0, 1)$ and $U = X + Y + V$. For $n \ge 1$ let $f_n = \mathbf{1}_{B_1\cup B_2}$, where $B_1 = \{x \in \{0,1\}^n : |x| \ge n/2\}$ and $B_2 = \{x : \frac{n}{4} + \frac{u_0\sqrt{n}}{4} \le |x| \le \frac{n}{4} + \frac{u_1\sqrt{n}}{4}\}$. Then $\operatorname{dist}(f_n) \to \frac{1}{2}$ and $\varepsilon(f_n) \to \varepsilon_\infty := \frac{1}{4} + P_U - 2P_3$, where*

$$P_U = \Pr[u_0 \le U \le u_1],$$
$$P_3 = \Pr[X \ge 0,\ Y \ge 0,\ u_0 \le U \le u_1] = \int_0^\infty q(s)\big(\Phi(u_1 - s) - \Phi(u_0 - s)\big)\, ds,$$

with $q(s) = \frac{e^{-s^2/4}\operatorname{erf}(s/2)}{2\sqrt{\pi}}$ and Φ the standard normal distribution function. Consequently $\limsup_{\varepsilon\to 0} \delta^*(\varepsilon)/\sqrt{\varepsilon}$ is at least $(4\varepsilon_\infty)^{-1/2}$. For $u_0 = 1.57$ and $u_1 = 12$, $\frac{1}{4\varepsilon_\infty} = 1.351168\dots$, so

$$\limsup_{\varepsilon\to 0} \frac{\delta^*(\varepsilon)}{\sqrt{\varepsilon}} \ > \ 1.1623.$$

Proof. Let $\mathbf{x}, \mathbf{y}$ be independent and uniform, and $X_n = (2|\mathbf{x}| - n)/\sqrt{n}$, $Y_n = (2|\mathbf{y}| - n)/\sqrt{n}$, $U_n = (4|\mathbf{x} \wedge \mathbf{y}| - n)/\sqrt{n}$. Then $(X_n, Y_n, U_n) = n^{-1/2} \sum_{i=1}^{n} \xi_i$, where the vectors $\xi_i = (2\mathbf{x}_i - 1,\ 2\mathbf{y}_i - 1,\ 4\mathbf{x}_i\mathbf{y}_i - 1)$ are i.i.d. with mean 0 and covariance matrix Σ with rows $(1,0,1)$, $(0,1,1)$, $(1,1,3)$ (e.g. $\mathbb{E}[(2\mathbf{x}_i - 1)(4\mathbf{x}_i\mathbf{y}_i - 1)] = 4\mathbb{E}[\mathbf{x}_i\mathbf{y}_i] - 0 = 1$, as $(2\mathbf{x}_i - 1)\mathbf{x}_i = \mathbf{x}_i$). This is the covariance matrix of (X, Y, U), and $\det \Sigma = 1$. By the multivariate central limit theorem, $(X_n, Y_n, U_n) \Rightarrow (X, Y, U)$. The sets $\{w_1 \geq 0\}$, $\{u_0 \leq w_3 \leq u_1\}$ and $\{w_1 \geq 0,\ w_2 \geq 0,\ u_0 \leq w_3 \leq u_1\}$ have boundaries contained in finitely many hyperplanes, which are null for the non-degenerate law of (X, Y, U); by the portmanteau theorem their probabilities converge.

Note that $\mathbf{x} \in B_1$ iff $X_n \geq 0$, and $\mathbf{x} \wedge \mathbf{y} \in B_2$ iff $u_0 \leq U_n \leq u_1$. If $\mathbf{x} \in B_2$ then $X_n \leq -\sqrt{n}/2 + u_1/2$, and if $\mathbf{x} \wedge \mathbf{y} \in B_1$ then $U_n \geq \sqrt{n}$; by Chebyshev's inequality ($\mathbb{E}X_n^2 = 1$, $\mathbb{E}U_n^2 = 3$) both have probability $o(1)$. Hence $\mu_n = \mathbb{E}f_n = \Pr[X_n \geq 0] + o(1) \to \frac{1}{2}$, $\Pr[f_n(\mathbf{x} \wedge \mathbf{y}) = 1] = \Pr[u_0 \leq U_n \leq u_1] + o(1) \to P_U$, and $\tau_n = \Pr[f_n(\mathbf{x}) = f_n(\mathbf{y}) = f_n(\mathbf{x} \wedge \mathbf{y}) = 1] = \Pr[X_n \geq 0, Y_n \geq 0, u_0 \leq U_n \leq u_1] + o(1) \to P_3$. Since $\mathbf{1}\{a \neq bc\} = a + bc - 2abc$ for bits, $\varepsilon(f_n) = \Pr[f_n(\mathbf{x} \wedge \mathbf{y}) = 1] + \mu_n^2 - 2\tau_n \to \varepsilon_\infty$.

For the distance, $\operatorname{dist}(f_n) \leq \Pr[f_n \neq 0] = \mu_n$, and $\Pr[f_n \neq 1] = 1 - \mu_n$. Let $S \neq \emptyset$, $s = |S|$, and $q_S = \Pr[f_n = 1 \mid \mathbf{x}_S = 1^S]$; then $\Pr[f_n \neq \mathrm{AND}_S] = \mu_n + 2^{-s}(1 - 2q_S)$. Given $\mathbf{x}_S = 1^S$, $|\mathbf{x}| = s + W$ with $W \sim \mathrm{Bin}(n - s, \frac{1}{2})$. For $n > s$ every point mass of W is at most $p := (n - s)^{-1/2}$ (proof of Theorem 1.2). By the symmetry of W about $\nu = \frac{n-s}{2}$, $\Pr[W \geq \nu] \leq \frac{1}{2} + \frac{p}{2}$, and the interval $[\nu - \frac{s}{2}, \nu)$ contains at most $\lceil s/2 \rceil$ integers, so $\Pr[s + W \geq \frac{n}{2}] = \Pr[W \geq \nu - \frac{s}{2}] \leq \frac{1}{2} + (\frac{s}{2} + 1)p$. Also $\Pr[s + W \leq \frac{n}{4} + \frac{u_1\sqrt{n}}{4}] \to 0$ for fixed s by Chebyshev's inequality. Fix $s_0 \geq 1$. For $s \leq s_0$ and $n > s_0$ this gives $2q_S - 1 \leq o(1)$ uniformly in S, so $\Pr[f_n \neq \mathrm{AND}_S] \geq \mu_n - 2^{-s}o(1) \geq \mu_n - o(1)$; for $s > s_0$, $\Pr[f_n \neq \mathrm{AND}_S] \geq \mu_n - 2^{-s_0}$. Hence $\liminf_n \operatorname{dist}(f_n) \geq \frac{1}{2} - 2^{-s_0}$ for every s_0, and $\operatorname{dist}(f_n) \to \frac{1}{2}$.

The density of $X + Y$ on the event $\{X \geq 0, Y \geq 0\}$ is $\int_0^s \varphi(x)\varphi(s - x)\,dx$, which equals $\frac{1}{2\pi}e^{-s^2/4}\int_0^s e^{-(x-s/2)^2}dx = q(s)$ for $s \geq 0$, where φ is the standard normal density; conditioning on $X + Y$ gives the integral for P_3. Since $\mathbf{1}\{a \neq bc\} = a + bc - 2abc$ and $\Pr[X \geq 0, Y \geq 0] = \frac{1}{4}$, $\varepsilon_\infty = \Pr\left[\mathbf{1}\{u_0 \leq U \leq u_1\} \neq \mathbf{1}\{X \geq 0,\ Y \geq 0\}\right] \geq \Pr[X \geq 0,\ Y \geq 0,\ U < u_0] > 0$, because (X, Y, U) has a positive density on $\mathbb{R}^3$. Hence $\operatorname{dist}(f_n)^2/\varepsilon(f_n) \to \frac{1}{4\varepsilon_\infty}$, so for every $r < \frac{1}{4\varepsilon_\infty}$ there is n with $\operatorname{dist}(f_n)^2 > r\,\varepsilon(f_n)$, and Lemma 7.1 gives $\delta^*(\varepsilon) > \sqrt{r\varepsilon}$ at $\varepsilon = 4^{-k}\varepsilon(f_n) \to 0$; this proves the general consequence.

For $u_0 = 1.57 = \frac{157}{100}$, $u_1 = 12$ the script `numerics/gaussian_certified.py` evaluates $P_U = \Phi(u_1/\sqrt{3}) - \Phi(u_0/\sqrt{3})$ and the integral of the non-negative integrand of P_3 over $[0, 30]$ in ball arithmetic (Arb, rigorous integration of analytic functions, 200 bits); u_0 enters as a ball of radius below 10^{-59} containing the rational $\frac{157}{100}$ (not as the nearest binary floating-point number). It obtains $P_U \in 0.1823515249974 \pm 10^{-13}$ and $\int_0^{30} \in 0.1236632380319 \pm 10^{-13}$; the tail is at most $\int_{30}^{\infty} e^{-s^2/4}/(2\sqrt{\pi})\,ds = \frac{1}{2}\mathrm{erfc}(15) < 10^{-99}$. So $\varepsilon_\infty \in 0.185025048933 \pm 10^{-12}$ and $\frac{1}{4\varepsilon_\infty} \in 1.35116840364 \pm 10^{-11}$, and $\sqrt{1.351168} > 1.1623$. □

Remark 7.5 (Even larger ratios). The bands of Proposition 7.3 were found by a local search over two-interval sets Z (floating point; `numerics/band_fast.py`). The best ratios found grow with n: 1.3017 ($n = 500$), 1.3153 (1000), 1.3260 (2000), 1.33335 (4000, exactly certified $> 4/3$), 1.3385 (8000, certified) and 1.3422 (16000, certified). Heuristically, as $n \to \infty$ the construction becomes Gaussian. Write $|\mathbf{x}| = \frac{n}{2} + \frac{\sqrt{n}}{2}X$, $|\mathbf{y}| = \frac{n}{2} + \frac{\sqrt{n}}{2}Y$ and $|\mathbf{x} \wedge \mathbf{y}| = \frac{n}{4} + \frac{\sqrt{n}}{4}(X + Y + V)$ with X, Y, V iid $N(0,1)$, put $f = g(X)$ near weight $n/2$ and $f = h(U)$ near weight $n/4$, and note that $\Pr[f \neq \mathrm{AND}_S] \to \mu + 2^{-|S|}(1 - 2\mu)$ for every fixed S. This gives $\operatorname{dist}^2/\varepsilon \to \min(\mu, 1 - \mu)^2/(\mu^2 - \Gamma)$, where $\Gamma = \mathbb{E}[(2\varphi(U) - 1)_+]$ and $\varphi(u) = \mathbb{E}[g(X)g(Y) \mid X + Y + V = u]$ for the optimal $h = \mathbf{1}\{\varphi > 1/2\}$. For $g = \mathbf{1}\{X \geq 0\}$ this is $1/(1 - 4\Gamma) \approx 1.35117$ ($\Gamma \approx 0.064975$), and neither a numerical optimisation over all g nor adding a few junta coordinates found anything better; the certified ratios above approach this value at rate roughly $n^{-1/2}$. For $g = \mathbf{1}\{X \geq 0\}$ the limit is made rigorous by

Proposition 7.4 (with h the indicator of $[u_0, u_1]$, $u_0 \approx 1.57$ the root of $\varphi = \frac{1}{2}$, and a cutoff u_1). We do not know the true value of $\limsup_{\varepsilon \to 0} \delta^*(\varepsilon)/\sqrt{\varepsilon} \in (1.1623, \sqrt{c_0}\,]$.

8 Approximate polymorphisms of an arbitrary Boolean function

The coupling of Lemma 3.1 is not specific to AND. In this section we use it to show that the dependence between ε and δ is polynomial for the approximate polymorphisms of *every* Boolean function, which answers Open Question 2 of Chase, Filmus, Minzer, Mossel and Saurabh [6].

Setting. Fix $m \geq 1$ and $g\colon \{0,1\}^m \to \{0,1\}$, and let $p_g = \mathbb{E}[g]$ (uniform measure). Let $\mathbf{Z}$ be a uniformly random $n \times m$ matrix over $\{0,1\}$ with rows $\mathbf{z}_1, \dots, \mathbf{z}_n \in \{0,1\}^m$ and columns $\mathbf{Z}^1, \dots, \mathbf{Z}^m \in \{0,1\}^n$, and put $g^n(\mathbf{Z}) = (g(\mathbf{z}_1), \dots, g(\mathbf{z}_n))$; note that $g^n(\mathbf{Z}) \sim \mu_{p_g}$. For a tuple $\mathbf{f} = (f_0, f_1, \dots, f_m)$ of functions $\{0,1\}^n \to \{0,1\}$ let

$$\mathrm{err}_g(\mathbf{f}) = \Pr\left[f_0(g^n(\mathbf{Z})) \neq g\big(f_1(\mathbf{Z}^1), \dots, f_m(\mathbf{Z}^m)\big)\right].$$

Following [6], $\mathbf{f}$ is a *δ-approximate multi-polymorphism* of g if $\mathrm{err}_g(\mathbf{f}) \leq \delta$, and an (exact) *multi-polymorphism* if $\mathrm{err}_g(\mathbf{f}) = 0$. A function f is a δ-approximate polymorphism if $(f, \dots, f)$ is a δ-approximate multi-polymorphism, and a pair (F_0, F_1) is a *skew polymorphism* if $(F_0, F_1, \dots, F_1)$ is a multi-polymorphism. Distances to f_0 are measured under μ_{p_g} and distances to $f_1, \dots, f_m$ under the uniform measure; "h is β-close to a constant" means $\Pr[h \neq b] \leq \beta$ for some $b \in \{0,1\}$, and otherwise h is *β-far from constants*. For $g = \mathrm{AND}_2$ and $f_0 = f_1 = f_2 = f$ we have $\mathrm{err}_g(\mathbf{f}) = \varepsilon(f)$, and the skew polymorphisms of AND_2 are exactly the pairs (h, h) with $h \in \mathcal{A}_n$ (put $y = x$ in $F_0(x \wedge y) = F_1(x)F_1(y)$ to get $F_0 = F_1$).

Chase et al. [6, Theorems 3.1 and 4.1] proved that for every g and $\varepsilon > 0$ there is $\delta = \delta(g, \varepsilon) > 0$ such that every δ-approximate polymorphism is ε-close to a skew polymorphism (their Theorem 3.1), and every δ-approximate multi-polymorphism is ε-close to a multi-polymorphism (their Theorem 4.1); Alekseev and Filmus [2] extended this to all predicates. Because of the regularity lemma, $\delta(g, \varepsilon)$ is of tower type in these proofs (doubly exponential with a decision-tree regularity lemma [6, Appendix C]). Polynomial dependence was known for affine g (linearity testing [4]), for the Arrow-type predicates by Keller [16, 17] (under restrictions on the functions in [16]; with the optimal dependence for the uniform distribution in [17]), and for Maj_3, where Chase et al. show linear dependence [6, Appendix D] (compare Theorem 8.2). Chase et al. asked: *"What is the optimal dependence between ε and δ in Theorem 3.1? Does it depend on g?"* [6, Open Question 2]. Alekseev and Filmus conjectured that for every predicate the optimal dependence is "polynomial or even linear" [2, Section 9, question 3].

Theorem 8.1 (Polynomial dependence for every g)**.** *Let $g\colon \{0,1\}^m \to \{0,1\}$ depend on exactly k of its coordinates. Put $c_g = 1$ if g is affine over $\mathbb{F}_2$ (this includes constant g), and otherwise*

$$c_g = \frac{1}{kK_k}, \qquad K_k = k(k-1)2^{k+2} + 3k - 2.$$

There is a constant C_g such that for every $n \geq 1$ and every tuple $\mathbf{f}$ with $\mathrm{err}_g(\mathbf{f}) = \delta$ there is a multi-polymorphism $\mathbf{F}$ of g with

$$\Pr_{\mu_{p_g}}[f_0 \neq F_0] \leq C_g\, \delta^{c_g} \qquad \text{and} \qquad \Pr[f_j \neq F_j] \leq C_g\, \delta^{c_g} \quad (j \in [m]).$$

Moreover, if $f_1 = \cdots = f_m$ then $\mathbf{F}$ can be chosen with $F_1 = \cdots = F_m$, and if in addition $p_g = 1/2$ and $f_0 = f_1$, then also with $F_0 = F_1$.

For balanced g the dependence is in fact linear; this is optimal (Proposition 8.16).

Theorem 8.2 (Linear dependence for balanced g)**.** *Let $g\colon \{0,1\}^m \to \{0,1\}$ with $p_g = 1/2$. There is a constant C'_g such that for every $n \geq 1$, every δ-approximate polymorphism f of g is $C'_g\delta$-close to a polymorphism of g.*

Corollary 8.3 (Answer to Open Question 2 of [6])**.** *Every δ-approximate polymorphism f of g is $C_g\delta^{c_g}$-close to a skew polymorphism (F_0, F_1) of g (that is, $\Pr_{\mu_{p_g}}[f \neq F_0] \leq C_g\delta^{c_g}$ and $\Pr[f \neq F_1] \leq C_g\delta^{c_g}$). Thus in [6, Theorems 3.1 and 4.1] one can take $\delta(g,\varepsilon) = (\varepsilon/C_g)^{1/c_g}$, a polynomial in ε. The optimal dependence does depend on g. Let $\delta_g(\varepsilon)$ be the supremum of the δ such that, for every n, every δ-approximate polymorphism of g is ε-close to a skew polymorphism. Then $\delta_g(\varepsilon) \geq \varepsilon/C'_g$ for every balanced g, and $\delta_g(\varepsilon) \leq C''_g\varepsilon$ for all small $\varepsilon > 0$ if moreover g depends on at least two coordinates (for instance XOR_m with $m \geq 2$, and Maj_3), whereas for every $m \geq 2$ and $0 < \varepsilon < 1/2$*

$$\frac{\varepsilon^m}{c_0(2m)^m} \;\leq\; \delta_{\mathrm{AND}_m}(\varepsilon) \;\leq\; \varepsilon^m, \qquad c_0 = 4 + 2\sqrt{2}.$$

In particular, every exponent $m = 1, 2, 3, \ldots$ occurs as the optimal exponent of some g. More generally, Theorem 8.18 below determines the optimal exponent for every non-constant unbalanced g, and Corollary 8.19 characterizes the g with linear loss.

Proof. The first two statements are the case $f_0 = \cdots = f_m = f$ of Theorem 8.1, and the balanced case is Theorem 8.2 and Proposition 8.16. Let $g = \mathrm{AND}_m$, so that $p_g = 2^{-m}$ and $g^n(\mathbf{Z}) = \mathbf{Z}^1 \wedge \cdots \wedge \mathbf{Z}^m$. If (F_0, F_1) is a skew polymorphism of AND_m, then taking all columns equal to x gives $F_0(x) = F_1(x)$. Taking $\mathbf{Z}^3 = \cdots = \mathbf{Z}^m = \mathbf{1}$ then gives $h(x \wedge y) = h(x)h(y)h(\mathbf{1})^{m-2}$ for $h = F_0 = F_1$. If $h(\mathbf{1}) = 0$ this gives $h(x) = h(x \wedge \mathbf{1}) = 0$ for all x (take $y = \mathbf{1}$); otherwise $h \in \mathcal{A}_n$. So the skew polymorphisms are exactly the pairs (h, h) with $h \in \mathcal{A}_n$.

Lower bound. Let f be a δ-approximate polymorphism of AND_m with $\delta = \varepsilon^m/(c_0(2m)^m)$, i.e. $\varepsilon_m(f) \leq \delta$ in the notation of Theorem 1.4(b). That theorem gives $h \in \mathcal{A}_n$ with $\Pr[f \neq h] \leq (c_0\delta)^{1/m} = \varepsilon/(2m)$. Since $h(x_1 \wedge \cdots \wedge x_m) = \prod_j h(x_j)$,

$$\Pr_{\mu_{2^{-m}}}[f \neq h] = \Pr[f(\mathbf{Z}^1 \wedge \cdots \wedge \mathbf{Z}^m) \neq h(\mathbf{Z}^1 \wedge \cdots \wedge \mathbf{Z}^m)] \leq \delta + \sum_{j=1}^{m} \Pr[f(\mathbf{Z}^j) \neq h(\mathbf{Z}^j)] \leq \delta + \frac{\varepsilon}{2} \leq \varepsilon.$$

So f is ε-close to the skew polymorphism (h, h).

Upper bound. Let $\delta > \varepsilon^m$. Since $\varepsilon^m < 2^{-m}$ we can pick δ' with $\varepsilon^m < \delta' < \min(\delta, 2^{-m})$. By Theorem 1.4(b), $\delta^*_m(\delta') \geq \delta'^{1/m} > \varepsilon$. So there is f with $\varepsilon_m(f) \leq \delta' < \delta$ and $\Pr[f \neq h] > \varepsilon$ for every $h \in \mathcal{A}_n$, and f is not ε-close to any skew polymorphism. □

8.1 Tools

For $j \in [m]$ and a *context* $u \in \{0,1\}^{[m]\setminus\{j\}}$ let $u^{j\to b} \in \{0,1\}^m$ be u with the bit b inserted in position j. The context u is *sensitive* for j if $g(u^{j\to 0}) \neq g(u^{j\to 1})$ and *insensitive* otherwise; let B_j and A_j be the sets of sensitive and insensitive contexts, and $s_j = |B_j|/2^{m-1}$, $a_j = |A_j|/2^{m-1} = 1 - s_j$. The function g depends on coordinate j iff $B_j \neq \emptyset$, and then $s_j \geq 2^{1-m}$. For $z \in \{0,1\}^m$ let $z_{-j} \in \{0,1\}^{[m]\setminus\{j\}}$ be its context.

Lemma 8.4 ([6, Lemma 3.2])**.** *If g depends on all its m coordinates and is not affine, then $a_j > 0$ for some $j \in [m]$.*

Proof. If $A_j = \emptyset$ for all j, then flipping any single coordinate of any input flips the value of g, so $g(z) = g(0) \oplus z_1 \oplus \cdots \oplus z_m$ is affine. □

For $T \subseteq [n]$, $q \in (0,1)$ and $h\colon \{0,1\}^n \to \{0,1\}$, the *rounding* of h to T is the T-junta $\mathrm{rd}_T^q h(x) = \mathbf{1}\{\mathbb{E}_{\mu_q}[h \mid \mathbf{x}_T = x_T] \geq 1/2\}$; we write $\mathrm{rd}_T = \mathrm{rd}_T^{1/2}$. Note that $\mathrm{rd}_T^q h$ depends only on h, q and T. The *majority value* of h under μ_q is the constant $\mathrm{rd}_\emptyset^q h$, i.e. 1 if $\Pr_{\mu_q}[h = 1] \geq 1/2$ and 0 otherwise; so ties are broken towards 1, here and below.

Lemma 8.5 (Rounding)**.** *Let $\mathbf{x} \sim \mu_q$ and let $\mathbf{x}'$ agree with $\mathbf{x}$ on T and be fresh outside T. Then*

$$\Pr_{\mu_q}[h \neq \mathrm{rd}_T^q h] \;\leq\; \Pr[h(\mathbf{x}) \neq h(\mathbf{x}')] \;=\; 2\sum_{S \not\subseteq T} \hat{h}(S)^2,$$

and $\Pr[h(\mathbf{x}) \neq h(\mathbf{x}')] \leq 2\Pr_{\mu_q}[h \neq J]$ for every T-junta J.

Proof. Given $\mathbf{x}_T = u$, let $r = \Pr[h(\mathbf{x}) = 1 \mid \mathbf{x}_T = u]$. Then $\Pr[h \neq \mathrm{rd}_T^q h \mid \mathbf{x}_T = u] = \min(r, 1-r) \leq 2r(1-r) = \Pr[h(\mathbf{x}) \neq h(\mathbf{x}') \mid \mathbf{x}_T = u]$. Average over u; the equality is Lemma 2.1. For the last claim note $J(\mathbf{x}) = J(\mathbf{x}')$, so $h(\mathbf{x}) \neq h(\mathbf{x}')$ forces $h(\mathbf{x}) \neq J(\mathbf{x})$ or $h(\mathbf{x}') \neq J(\mathbf{x}')$, and $\mathbf{x}' \sim \mu_q$. □

Lemma 8.6 (Noise)**.** *Let $q \in (0,1)$ and $\rho \in [0,1)$. Let $(\mathbf{y}, \mathbf{y}')$ be a random pair in $\{0,1\}^N \times \{0,1\}^N$ whose coordinate pairs $(\mathbf{y}_i, \mathbf{y}'_i)$ are i.i.d. and exchangeable, with $\Pr[\mathbf{y}_i = 1] = q$ and $\Pr[\mathbf{y}_i \neq \mathbf{y}'_i] = 2q(1-q)(1-\rho)$. Then for every $h\colon \{0,1\}^N \to \{0,1\}$, with b the majority value of h under μ_q,*

$$\Pr_{\mu_q}[h \neq b] \leq \frac{\Pr[h(\mathbf{y}) \neq h(\mathbf{y}')]}{1-\rho}.$$

Proof. An exchangeable pair of Bernoulli(q) bits is determined by the probability that the bits differ, so $(\mathbf{y}_i, \mathbf{y}'_i)$ has the law of "$\mathbf{y}'_i = \mathbf{y}_i$ with probability ρ, and $\mathbf{y}'_i$ is a fresh Bernoulli(q) bit otherwise". Hence $\mathbb{E}[\phi_S(\mathbf{y})\phi_T(\mathbf{y}')] = \rho^{|S|}\mathbf{1}\{S = T\}$ and $\mathbb{E}[h(\mathbf{y})h(\mathbf{y}')] = \sum_S \rho^{|S|}\hat{h}(S)^2 \leq \hat{h}(\emptyset)^2 + \rho\sum_{S\neq\emptyset}\hat{h}(S)^2$, using $\rho \geq 0$. With $\nu = \mathbb{E}h$ we get $\Pr[h(\mathbf{y}) \neq h(\mathbf{y}')] = 2\nu - 2\mathbb{E}[h(\mathbf{y})h(\mathbf{y}')] \geq 2(1-\rho)\sum_{S\neq\emptyset}\hat{h}(S)^2 = 2(1-\rho)\nu(1-\nu) \geq (1-\rho)\min(\nu, 1-\nu)$, and $\min(\nu, 1-\nu) = \Pr[h \neq b]$. □

Lemma 8.7 (One resampled coordinate)**.** *Let g depend on coordinate j, let $\mathbf{z}$ be uniform on $\{0,1\}^m$, and let $\mathbf{z}'$ be $\mathbf{z}$ with coordinate j resampled. Then $(g(\mathbf{z}), g(\mathbf{z}'))$ is an exchangeable pair of Bernoulli(p_g) bits with $\Pr[g(\mathbf{z}) \neq g(\mathbf{z}')] = s_j/2 = 2p_g(1-p_g)(1-\rho_j)$ for some $\rho_j \in [0,1)$, and $\frac{1}{1-\rho_j} = \frac{4p_g(1-p_g)}{s_j} \leq 2^{m-1}$.*

Proof. Exchangeability and the marginals are clear, and $g(\mathbf{z}) \neq g(\mathbf{z}')$ iff $\mathbf{z}_{-j} \in B_j$ and $\mathbf{z}_j \neq \mathbf{z}'_j$, which has probability $s_j/2 > 0$. Let $r(u) = \Pr[g(\mathbf{z}) = 1 \mid \mathbf{z}_{-j} = u]$. Given $\mathbf{z}_{-j}$, the bits $g(\mathbf{z}), g(\mathbf{z}')$ are independent, so $\Pr[g(\mathbf{z}) = g(\mathbf{z}') = 1] = \mathbb{E}[r(\mathbf{z}_{-j})^2] \geq p_g^2$. Therefore $s_j/2 = 2p_g - 2\Pr[g(\mathbf{z}) = g(\mathbf{z}') = 1] \leq 2p_g(1-p_g)$, i.e. $\rho_j \geq 0$; and $\rho_j < 1$ since $s_j > 0$. Finally $4p_g(1-p_g) \leq 1$ and $s_j \geq 2^{1-m}$. □

8.2 The junta lemma

For $j \in [m]$ let $\mathbf{c}_j = (f_k(\mathbf{Z}^k))_{k\neq j} \in \{0,1\}^{[m]\setminus\{j\}}$ and $\mathbf{R}_j = \{i \in [n] : (\mathbf{z}_i)_{-j} \in A_j\}$, the set of rows whose context is insensitive for j. Both are functions of the columns $\mathbf{Z}^k$, $k \neq j$. Write $\delta = \mathrm{err}_g(\mathbf{f})$.

Lemma 8.8 (Spectral containment for general g)**.** *For every $j \in [m]$, with $\hat{f}_j$ the uniform-measure Fourier coefficients,*

$$\sum_{S\subseteq[n]} \hat{f}_j(S)^2 \Pr\left[\mathbf{c}_j \in B_j,\; S \cap \mathbf{R}_j \neq \emptyset\right] \;\leq\; \delta.$$

Proof. Let $\mathbf{W}$ be obtained from $\mathbf{Z}$ by replacing the entries $\mathbf{Z}_{ij}$, $i \in \mathbf{R}_j$, by fresh independent uniform bits. Since $\mathbf{R}_j$ is determined by the columns $\mathbf{Z}^k$, $k \neq j$, and $\mathbf{Z}^j$ is independent of them, $\mathbf{W}$ is uniform. Its columns $k \neq j$ are those of $\mathbf{Z}$, so $\mathbf{c}_j(\mathbf{W}) = \mathbf{c}_j(\mathbf{Z})$. Moreover $g^n(\mathbf{W}) = g^n(\mathbf{Z})$: a row $i \notin \mathbf{R}_j$ is unchanged, and a row $i \in \mathbf{R}_j$ changes only in coordinate j, while its context is insensitive. Let E be the event that $\mathbf{c}_j \in B_j$ and $f_j(\mathbf{Z}^j) \neq f_j(\mathbf{W}^j)$. On E the two right-hand sides $g(\mathbf{c}_j^{j\to f_j(\mathbf{Z}^j)})$ and $g(\mathbf{c}_j^{j\to f_j(\mathbf{W}^j)})$ differ, while the left-hand sides $f_0(g^n(\mathbf{Z})) = f_0(g^n(\mathbf{W}))$ agree, so the test fails on $\mathbf{Z}$ or on $\mathbf{W}$; hence $\Pr[E] \le 2\delta$. On the other hand, conditionally on the columns $k \neq j$, the pair $(\mathbf{Z}^j, \mathbf{W}^j)$ is the pair of Lemma 2.1 (with $p = 1/2$) for $J = [n] \setminus \mathbf{R}_j$, so

$$\Pr[E] = \mathbb{E}\Big[\mathbf{1}\{\mathbf{c}_j \in B_j\} \cdot 2 \sum_{S \cap \mathbf{R}_j \neq \emptyset} \hat{f}_j(S)^2\Big] = 2 \sum_S \hat{f}_j(S)^2 \Pr[\mathbf{c}_j \in B_j,\ S \cap \mathbf{R}_j \neq \emptyset]. \qquad \square$$

Lemma 8.9 (Junta lemma)**.** *Let $j \in [m]$ with $a_j > 0$ and $P_j := \Pr[\mathbf{c}_j \in B_j] > 0$, and let $T_j = \{i \in [n] : \Pr[i \in \mathbf{R}_j \mid \mathbf{c}_j \in B_j] \le a_j/4\}$. Then*

$$(a)\ \ |T_j| \le \frac{8}{a_j} \ln \frac{2}{P_j}, \qquad (b)\ \ \Pr[f_j \neq \mathrm{rd}_{T_j} f_j] \le 2 \sum_{S \not\subseteq T_j} \hat{f}_j(S)^2 \le \frac{8\delta}{a_j P_j}.$$

Proof. (b) If $S \not\subseteq T_j$, pick $i \in S \setminus T_j$; then $\Pr[\mathbf{c}_j \in B_j,\ S \cap \mathbf{R}_j \neq \emptyset] \ge \Pr[\mathbf{c}_j \in B_j,\ i \in \mathbf{R}_j] > \frac{a_j}{4} P_j$. Lemma 8.8 gives $\sum_{S \not\subseteq T_j} \hat{f}_j(S)^2 \le 4\delta/(a_j P_j)$, and Lemma 8.5 gives the first inequality.

(a) We may assume $T_j \neq \emptyset$. The rows are independent and uniform, so the events $\{i \in \mathbf{R}_j\}$, $i \in [n]$, are independent, each of probability a_j; hence $\mathbf{N} = |\mathbf{R}_j \cap T_j|$ is binomial with mean $a_j|T_j|$, and the Chernoff bound gives $\Pr[\mathbf{N} \le a_j|T_j|/2] \le e^{-a_j|T_j|/8}$. On the other hand $\mathbb{E}[\mathbf{N} \mid \mathbf{c}_j \in B_j] \le a_j|T_j|/4$ by the definition of T_j, so by Markov's inequality $\Pr[\mathbf{N} < a_j|T_j|/2 \mid \mathbf{c}_j \in B_j] \ge 1/2$. Thus $P_j/2 \le \Pr[\mathbf{N} < a_j|T_j|/2] \le e^{-a_j|T_j|/8}$, which is (a). $\square$

8.3 Propagation, certificates and exactness

Lemma 8.10 (From one column to f_0)**.** *Let g depend on coordinate j, let $T \subseteq [n]$, and suppose $\Pr[f_j \neq J] \le \beta$ for some T-junta J. Then*

$$\Pr_{\mu_{p_g}}[f_0 \neq \mathrm{rd}_T^{p_g} f_0] \ \le\ \frac{2\delta + 2\beta}{1 - \rho_j} \ \le\ 2^m(\delta + \beta).$$

Proof. Let $\mathbf{W}$ be obtained from $\mathbf{Z}$ by resampling the entries $\mathbf{Z}_{ij}$, $i \notin T$. Then $\mathbf{W}$ is uniform, its columns $k \neq j$ are those of $\mathbf{Z}$, and $\Pr[f_j(\mathbf{Z}^j) \neq f_j(\mathbf{W}^j)] \le 2\beta$ by the last part of Lemma 8.5. Hence the right-hand sides of the test on $\mathbf{Z}$ and on $\mathbf{W}$ differ with probability at most 2β, and $\mathbf{y} = g^n(\mathbf{Z})$, $\mathbf{y}' = g^n(\mathbf{W})$ satisfy $\Pr[f_0(\mathbf{y}) \neq f_0(\mathbf{y}')] \le 2\delta + 2\beta$. Now $\mathbf{y}_T = \mathbf{y}'_T$, and the pairs $(\mathbf{y}_i, \mathbf{y}'_i)$, $i \notin T$, are i.i.d., independent of $\mathbf{y}_T$, with the law of Lemma 8.7. Given $\mathbf{y}_T = u$, Lemma 8.6 (with $q = p_g$, $\rho = \rho_j$) applied to $h = f_0(u, \cdot)$ shows that h differs from its majority value, which is $\mathrm{rd}_T^{p_g} f_0(u)$, with probability at most $\Pr[f_0(\mathbf{y}) \neq f_0(\mathbf{y}') \mid \mathbf{y}_T = u]/(1 - \rho_j)$. Averaging over u and using Lemma 8.7 gives the claim. $\square$

Lemma 8.11 (From f_0 to every column)**.** *Let $T \subseteq [n]$, suppose $\Pr_{\mu_{p_g}}[f_0 \neq J_0] \le \beta$ for some T-junta J_0, let g depend on coordinate j, and suppose that f_k is γ-far from constants for every $k \in [m] \setminus \{j\}$. Then $\Pr[f_j \neq \mathrm{rd}_T f_j] \le (2\delta + 2\beta)\,\gamma^{-2(m-1)}$.*

Proof. Let $\mathbf{W}$ be obtained from $\mathbf{Z}$ by resampling all rows $i \notin T$. Then $\mathbf{W}$ is uniform, $g^n(\mathbf{W})$ agrees with $g^n(\mathbf{Z})$ on T, and both have law μ_{p_g}, so $f_0(g^n(\mathbf{Z})) \neq f_0(g^n(\mathbf{W}))$ with probability at

most 2β (as in Lemma 8.5). Hence the right-hand sides $\mathrm{RHS}(\mathbf{Z})$, $\mathrm{RHS}(\mathbf{W})$ of the test differ with probability at most $2\delta + 2\beta$. Fix $v \in B_j$ and let E be the event that $f_k(\mathbf{Z}^k) = f_k(\mathbf{W}^k) = v_k$ for all $k \neq j$. The column pairs $(\mathbf{Z}^k, \mathbf{W}^k)$, $k \in [m]$, are independent, and each is the pair of Lemma 2.1 with $J = T$. With $r_k(u) = \Pr[f_k(\mathbf{x}) = v_k \mid \mathbf{x}_T = u]$ for uniform $\mathbf{x}$, we get $\Pr[f_k(\mathbf{Z}^k) = f_k(\mathbf{W}^k) = v_k] = \mathbb{E}[r_k^2] \geq (\mathbb{E} r_k)^2 \geq \gamma^2$, so $\Pr[E] \geq \gamma^{2(m-1)}$. On E we have $\mathrm{RHS}(\mathbf{Z}) = g(v^{j \to f_j(\mathbf{Z}^j)})$ and $\mathrm{RHS}(\mathbf{W}) = g(v^{j \to f_j(\mathbf{W}^j)})$; as v is sensitive, they differ iff $f_j(\mathbf{Z}^j) \neq f_j(\mathbf{W}^j)$. Since E is independent of $(\mathbf{Z}^j, \mathbf{W}^j)$,

$$\Pr[f_j(\mathbf{Z}^j) \neq f_j(\mathbf{W}^j)] = \Pr[\mathrm{RHS}(\mathbf{Z}) \neq \mathrm{RHS}(\mathbf{W}) \mid E] \leq \frac{2\delta + 2\beta}{\gamma^{2(m-1)}},$$

and Lemma 8.5 finishes the proof. □

Lemma 8.12 (Certificates)**.** *Let g depend on all its m coordinates, $0 < \varepsilon \leq 1/2$, $\varepsilon' = 2^{-2m-2}\varepsilon^m$ and $\delta \leq \varepsilon'$. Suppose that some f_{k_0}, $k_0 \in [m]$, is ε'-close to a constant. Let J be the set of $j \in [m]$ such that f_j is ε-close to a constant, let a_j be the majority value of f_j ($j \in J$), and let a_0 be the majority value of f_0 under μ_{p_g}. Then $\Pr_{\mu_{p_g}}[f_0 \neq a_0] \leq \varepsilon$, and $(F_0, \ldots, F_m)$ defined by $F_0 \equiv a_0$, $F_j \equiv a_j$ ($j \in J$) and $F_j = f_j$ ($j \notin J$) is a multi-polymorphism of g. All its distances to $\mathbf{f}$ are at most ε.*

Proof. A constant is a $\emptyset$-junta, so Lemma 8.10 with $T = \emptyset$ and $j = k_0$ gives $\Pr[f_0 \neq a_0] \leq 2^m(\delta + \varepsilon') \leq 2^{m+1}\varepsilon' \leq \varepsilon$. Fix $x \in \{0,1\}^m$ with $x_J = a_J$. For $j \in J$, $\Pr[f_j = x_j] \geq 1 - \varepsilon \geq \frac{1}{2}$; for $j \notin J$, f_j is ε-far from constants, so $\Pr[f_j = x_j] > \varepsilon$. The columns are independent and $|J| \geq 1$, so $\mathrm{RHS}(\mathbf{Z}) = g(x)$ with probability at least $2^{-|J|}\varepsilon^{m-|J|} \geq 2^{-m}\varepsilon^{m-1}$. On the other hand $\mathrm{RHS}(\mathbf{Z}) \neq a_0$ with probability at most $\delta + \Pr[f_0(g^n(\mathbf{Z})) \neq a_0] \leq 2^{m+2}\varepsilon' = 2^{-m}\varepsilon^m < 2^{-m}\varepsilon^{m-1}$. Hence there is a $\mathbf{Z}$ with $\mathrm{RHS}(\mathbf{Z}) = g(x)$ and $\mathrm{RHS}(\mathbf{Z}) = a_0$, i.e. $g(x) = a_0$. So $x_J = a_J$ forces $g(x) = a_0$, which gives $F_0(g^n(\mathbf{Z})) = a_0 = g(F_1(\mathbf{Z}^1), \ldots, F_m(\mathbf{Z}^m))$ for all $\mathbf{Z}$. Finally $\Pr[f_j \neq a_j] \leq \varepsilon$ for $j \in J$, because a_j is the majority value of a function that is ε-close to a constant. □

Lemma 8.13 (Exactness on a junta)**.** *Let $T \subseteq [n]$, and let $F_0, \ldots, F_m$ be T-juntas with $\delta + \Pr_{\mu_{p_g}}[f_0 \neq F_0] + \sum_{j=1}^m \Pr[f_j \neq F_j] < 2^{-m|T|}$. Then $\mathbf{F}$ is a multi-polymorphism of g.*

Proof. By the union bound, $\mathrm{err}_g(\mathbf{F}) < 2^{-m|T|}$. Whether $F_0(g^n(\mathbf{Z})) = g(F_1(\mathbf{Z}^1), \ldots, F_m(\mathbf{Z}^m))$ holds depends only on the rows $\mathbf{z}_i$, $i \in T$. Each of the $2^{m|T|}$ possible values of these rows has probability $2^{-m|T|}$, so the identity holds for all of them, i.e. for every $\mathbf{Z}$. □

Lemma 8.14 (Affine g)**.** *Let $g(x) = x_1 \oplus \cdots \oplus x_m \oplus b$ with $m \geq 1$, and let $\delta = \mathrm{err}_g(\mathbf{f}) < \min(1/50, 1/(m+2))$. Then there is a multi-polymorphism $\mathbf{F}$ with all distances at most δ, such that $F_\ell = F_{\ell'}$ whenever $f_\ell = f_{\ell'}$ ($\ell, \ell' \in \{0, \ldots, m\}$).*

Proof. Here $p_g = 1/2$ and $g^n(\mathbf{Z}) = \mathbf{Z}^1 \oplus \cdots \oplus \mathbf{Z}^m \oplus b\mathbf{1}$, where $\mathbf{1}$ is the all-ones vector.

$m = 1$. If $f_0 \neq f_1$, take $F_1 = f_1$ and $F_0(y) = f_1(y \oplus b\mathbf{1}) \oplus b$; this is exact, and $\Pr[f_0 \neq F_0] = \delta$ because $y = \mathbf{Z}^1 \oplus b\mathbf{1}$ is uniform. If $f_0 = f_1 = f$ and $b = 0$, then $\delta = 0$ and $\mathbf{F} = \mathbf{f}$. If $f_0 = f_1 = f$ and $b = 1$, let $F(x) = f(x)$ if $x_1 = 0$ and $F(x) = f(x \oplus \mathbf{1}) \oplus 1$ if $x_1 = 1$. Then $F(x \oplus \mathbf{1}) = F(x) \oplus 1$ for all x, so (F, F) is exact. The event $D = \{f(x) \neq f(x \oplus \mathbf{1}) \oplus 1\}$ has probability δ and is invariant under $x \mapsto x \oplus \mathbf{1}$, which exchanges $\{x_1 = 0\}$ and $\{x_1 = 1\}$. Hence $\Pr[f \neq F] = \frac{1}{2}\Pr[D \mid x_1 = 1] = \delta/2$.

$m \geq 2$. Let $\Phi_\ell = (-1)^{f_\ell}$, $\chi_S(x) = (-1)^{\sum_{i \in S} x_i}$ and $\hat{\Phi}_\ell(S) = \mathbb{E}[\Phi_\ell \chi_S]$ (uniform measure). Since $\chi_S(g^n(\mathbf{Z})) = (-1)^{b|S|} \prod_k \chi_S(\mathbf{Z}^k)$ and the columns are independent,

$$1 - 2\delta = \mathbb{E}\Big[\Phi_0(g^n(\mathbf{Z}))\,(-1)^b \prod_{k=1}^m \Phi_k(\mathbf{Z}^k)\Big] = \sum_S (-1)^{b(|S|+1)} \prod_{\ell=0}^m \hat{\Phi}_\ell(S).$$

Given $\ell \in \{0,\dots,m\}$ choose two further indices ℓ_1, ℓ_2 (possible as $m+1 \geq 3$). Bounding the remaining factors by 1 and using Cauchy–Schwarz and Parseval,

$$1 - 2\delta \leq \max_S |\hat{\Phi}_\ell(S)| \cdot \sum_S |\hat{\Phi}_{\ell_1}(S)\hat{\Phi}_{\ell_2}(S)| \leq \max_S |\hat{\Phi}_\ell(S)|.$$

So there is S_ℓ with $|\hat{\Phi}_\ell(S_\ell)| \geq 1-2\delta$, and then $\hat{\Phi}_\ell(S)^2 \leq 1-(1-2\delta)^2 \leq 4\delta$ for $S \neq S_\ell$. If $S_\ell \neq S_{\ell'}$ for some $\ell \neq \ell'$, then

$$1 - 2\delta \leq \sum_S |\hat{\Phi}_\ell(S)\hat{\Phi}_{\ell'}(S)| \leq |\hat{\Phi}_{\ell'}(S_\ell)| + |\hat{\Phi}_\ell(S_{\ell'})| + \Big(\sum_{S\neq S_\ell} \hat{\Phi}_\ell(S)^2\Big)^{\frac{1}{2}} \Big(\sum_{S\neq S_{\ell'}} \hat{\Phi}_{\ell'}(S)^2\Big)^{\frac{1}{2}} \leq 4\sqrt{\delta} + 4\delta,$$

which is false for $\delta < 1/50$. Hence all S_ℓ equal some S. Let $(-1)^{a_\ell}$ be the sign of $\hat{\Phi}_\ell(S)$ and $F_\ell(x) = \bigoplus_{i\in S} x_i \oplus a_\ell$. Then $\Pr[f_\ell \neq F_\ell] = (1-|\hat{\Phi}_\ell(S)|)/2 \leq \delta$ (for $\ell = 0$ the relevant measure is $\mu_{p_g} = \mu_{1/2}$), and $F_\ell = F_{\ell'}$ when $f_\ell = f_{\ell'}$. Finally, $F_0(g^n(\mathbf{Z})) \oplus g(F_1(\mathbf{Z}^1),\dots,F_m(\mathbf{Z}^m)) = b|S| \oplus a_0 \oplus a_1 \oplus \cdots \oplus a_m \oplus b$ is a constant, so $\mathrm{err}_g(\mathbf{F}) \in \{0,1\}$. As $\mathrm{err}_g(\mathbf{F}) \leq \delta + (m+1)\delta < 1$, $\mathbf{F}$ is exact. □

8.4 Proof of Theorem 8.1

Trivial range. If $\delta = 0$ then $\mathbf{F} = \mathbf{f}$ is a multi-polymorphism. Dictators $F_0 = \cdots = F_m = x_1$ form a multi-polymorphism of every g (both sides equal $g(\mathbf{z}_1)$) and satisfy all the equalities in the "moreover" part. So it suffices to prove the theorem for $0 < \delta < \delta_g$, for some $\delta_g > 0$, with $C_g \geq \delta_g^{-c_g}$.

Reduction to $k = m$. Let $K \subseteq [m]$ be the set of coordinates on which g depends, and $G\colon \{0,1\}^K \to \{0,1\}$ the induced function. Then $p_G = p_g$, $\mathrm{err}_g(\mathbf{f}) = \mathrm{err}_G(f_0, (f_j)_{j\in K})$, and if $(F_0, (F_j)_{j\in K})$ is a multi-polymorphism of G then extending it by arbitrary F_j, $j \notin K$, gives a multi-polymorphism of g. Take $F_j = f_j$ for $j \notin K$ in general. If $f_1 = \cdots = f_m$, the functions F_j, $j \in K$, are equal (by the theorem for G), and we take F_j equal to them for $j \notin K$ too. If $k = 0$ then $g \equiv b$, $g^n(\mathbf{Z}) = b\mathbf{1}$ and $\mathrm{err}_g(\mathbf{f}) = \mathbf{1}\{f_0(b\mathbf{1}) \neq b\}$, so $\delta < 1$ means that $\mathbf{f}$ is exact. So assume that g depends on all $m = k \geq 1$ coordinates.

Affine g. Lemma 8.14 applies for $\delta < \min(1/50, 1/(m+2))$, with distances at most δ.

Non-affine g. Then $m \geq 2$ (functions of one variable are affine), and by Lemma 8.4 we may fix $j^* \in [m]$ with $a_{j^*} > 0$; note $a_{j^*} \geq 2^{1-m}$. Let $K = K_m$ and

$$\varepsilon_0 = \frac{e^{-m2^{m+2}}}{(m+2)\,2^{3m+6}}.$$

Given δ, define ε' and ε by $\delta = \varepsilon'^K$ and $\varepsilon' = 2^{-2m-2}\varepsilon^m$, so that $\varepsilon = 2^{(2m+2)/m}\delta^{1/(mK)} \leq 8\,\delta^{1/(mK)}$. We assume $\varepsilon \leq \varepsilon_0$, which holds when δ is small enough; then $\delta \leq \varepsilon' \leq \varepsilon_0$.

Case I: some f_k, $k \in [m]$, is ε'-close to a constant. Lemma 8.12 gives a multi-polymorphism within distance ε.

Case II: every f_k, $k \in [m]$, is ε'-far from constants. Fix $v \in B_{j^*}$. By independence of the columns, $P_{j^*} \geq \Pr[\mathbf{c}_{j^*} = v] = \prod_{k\neq j^*} \Pr[f_k = v_k] \geq \varepsilon'^{\,m-1}$. Let $T = T_{j^*}$ be the set of Lemma 8.9; then

$$|T| \leq 2^{m+2}\ln\big(2\varepsilon'^{-(m-1)}\big), \qquad \beta_{j^*} := \Pr[f_{j^*} \neq \mathrm{rd}_T f_{j^*}] \leq 2^{m+2}\delta\,\varepsilon'^{-(m-1)}.$$

Lemma 8.10 (with $J = \mathrm{rd}_T f_{j^*}$) gives $\beta_0 := \Pr_{\mu_{p_g}}[f_0 \neq \mathrm{rd}_T^{p_g} f_0] \leq 2^m(\delta + \beta_{j^*}) \leq 2^{2m+3}\delta\,\varepsilon'^{-(m-1)}$. Lemma 8.11 (with $J_0 = \mathrm{rd}_T^{p_g} f_0$ and $\gamma = \varepsilon'$) gives, for every $j \in [m]$,

$$\beta_j := \Pr[f_j \neq \mathrm{rd}_T f_j] \leq 2(\delta + \beta_0)\,\varepsilon'^{-2(m-1)} \leq 2^{2m+5}\delta\,\varepsilon'^{-3(m-1)}.$$

Put $D = 2^{3m+6}\delta\,\varepsilon'^{-3(m-1)} = 2^{3m+6}\varepsilon'^{\,m(m-1)2^{m+2}+1}$, so that δ and all β_ℓ are at most D. Since $\ln 2 < 1$, $m|T|\ln 2 \le m2^{m+2}+m(m-1)2^{m+2}\ln(1/\varepsilon')$, i.e. $2^{-m|T|} \ge e^{-m2^{m+2}}\varepsilon'^{\,m(m-1)2^{m+2}}$. Therefore

$$\delta + \beta_0 + \sum_{j=1}^{m}\beta_j \le (m+2)D = (m+2)2^{3m+6}\varepsilon'\cdot\varepsilon'^{\,m(m-1)2^{m+2}} < e^{-m2^{m+2}}\varepsilon'^{\,m(m-1)2^{m+2}} \le 2^{-m|T|},$$

because $\varepsilon' < \varepsilon_0$. By Lemma 8.13, $\mathbf{F} = (\mathrm{rd}_T^{p_g} f_0, \mathrm{rd}_T f_1, \dots, \mathrm{rd}_T f_m)$ is a multi-polymorphism of g, and its distances to $\mathbf{f}$ are at most $D \le 2^{3m+6}\varepsilon'^{\,2} \le \varepsilon' \le \varepsilon$.

In both cases all distances are at most $\varepsilon \le 8\delta^{1/(mK)} = 8\delta^{c_g}$, which proves the bound (with δ_g chosen so that $\varepsilon \le \varepsilon_0$ for $\delta < \delta_g$, and $C_g = \max(8, \delta_g^{-c_g})$).

The "moreover" part. In Case II, $\mathrm{rd}_T f_j$ depends only on f_j, so $f_1 = \cdots = f_m$ gives $F_1 = \cdots = F_m$; if moreover $p_g = 1/2$ and $f_0 = f_1$, then $\mathrm{rd}_T^{p_g} f_0 = \mathrm{rd}_T f_1$. In Case I, if $f_1 = \cdots = f_m = f$, then $k_0 \in J$ forces $J = [m]$, and all a_j equal the majority value a of f; if moreover $p_g = 1/2$ and $f_0 = f$, then $a_0 = a$ (both are the majority value of f under the uniform measure). The affine case is covered by Lemma 8.14, and the reduction step preserves these equalities. □

8.5 Balanced g: proof of Theorem 8.2

The polynomial loss in Theorem 8.1 comes only from functions that are close to constants: if all f_k are γ-far from constants for a fixed γ, the proof of Case II gives a junta of bounded size with linear error. For a single function and balanced g, functions close to constants are handled by small-set expansion. For $\rho \in (-1,1)$, a pair $(\mathbf{x},\mathbf{y})$ of uniform points of $\{0,1\}^n$ is *ρ-correlated* if the pairs $(\mathbf{x}_i,\mathbf{y}_i)$ are i.i.d. with $\Pr[\mathbf{y}_i = \mathbf{x}_i] = (1+\rho)/2$.

Lemma 8.15 (Small-set expansion)**.** *If $(\mathbf{x},\mathbf{y})$ is ρ-correlated and $A \subseteq \{0,1\}^n$ has $|A| = \eta 2^n$, then* $\Pr[\mathbf{x}\in A,\ \mathbf{y}\in A] \le \eta^{2/(1+|\rho|)}$.

Proof. First let $\rho \ge 0$ and let $B \subseteq \{0,1\}^n$ be any set with $|B| = \eta 2^n$. With T_σ the noise operator and $\|\cdot\|_q$ the $L^q(\mu_{1/2})$ norms, $\Pr[\mathbf{x}\in A,\mathbf{y}\in B] = \langle \mathbf{1}_A, \mathrm{T}_\rho \mathbf{1}_B\rangle = \langle \mathrm{T}_{\sqrt{\rho}}\mathbf{1}_A, \mathrm{T}_{\sqrt{\rho}}\mathbf{1}_B\rangle$. By Cauchy–Schwarz and the hypercontractive inequality $\|\mathrm{T}_{\sqrt{\rho}}h\|_2 \le \|h\|_{1+\rho}$ of Bonami [5] (see [25, Chapter 9]), this is at most $\|\mathbf{1}_A\|_{1+\rho}\|\mathbf{1}_B\|_{1+\rho} = \eta^{2/(1+\rho)}$. Taking $B = A$ proves the case $\rho \ge 0$. If $\rho < 0$, then $(\mathbf{x}, \mathbf{1}\oplus\mathbf{y})$ is $|\rho|$-correlated and $\Pr[\mathbf{x}\in A,\mathbf{y}\in A] = \Pr[\mathbf{x}\in A,\ \mathbf{1}\oplus\mathbf{y}\in B]$ with $B = \mathbf{1}\oplus A$, so the two-set bound applies. □

Proof of Theorem 8.2. Let $f_0 = \cdots = f_m = f$ and $\delta = \mathrm{err}_g(\mathbf{f})$. Dictators are polymorphisms, so it suffices to treat $\delta < \delta_g'$ for a constant $\delta_g' \in (0,1/2]$ chosen below, with $C_g' \ge 1/\delta_g'$. If g depends only on the coordinates in $K \subsetneq [m]$ and G is the induced function, then $p_G = p_g$, $\mathrm{err}_g(f,\dots,f) = \mathrm{err}_G(f,\dots,f)$, and g and G have the same polymorphisms (as $g^n(\mathbf{Z}) = G^n((\mathbf{Z}^j)_{j\in K})$); and $K \ne \emptyset$ because $p_g = 1/2$. So we may assume that g depends on all m coordinates. If g is affine, we take $\delta_g' = \min\{1/50, 1/(m+2)\}$ and $C_g' = 1/\delta_g'$, and Lemma 8.14 applies (all F_ℓ coincide since all f_ℓ do). So let g be non-affine; then $m \ge 2$, and we fix j^* with $a_{j^*} \ge 2^{1-m}$ (Lemma 8.4).

For $j \in [m]$ the pairs $(g(\mathbf{z}_i), \mathbf{z}_{ij})$, $i \in [n]$, are i.i.d. pairs of uniform bits. A pair of uniform bits is determined by the probability that they agree, so $(g^n(\mathbf{Z}), \mathbf{Z}^j)$ is ρ_j-correlated with $\rho_j = 2\Pr[g(\mathbf{z}) = \mathbf{z}_j] - 1$. As g depends on at least two coordinates, $g \notin \{x_j, 1\oplus x_j\}$, so $0 < \Pr[g(\mathbf{z}) = \mathbf{z}_j] < 1$. Since this probability is a multiple of 2^{-m}, $r := \max_j|\rho_j| \le 1 - 2^{1-m}$. Put

$$\gamma = \min\Big\{\frac{1}{2(m+1)},\ (2m)^{-\frac{1+r}{1-r}}\Big\}.$$

Case A: f is γ-close to a constant. Let b be the majority value of f, $A = \{x : f(x) \neq b\}$ and $\eta = \Pr[f \neq b] \leq \gamma$. If $g(b,\dots,b) \neq b$, then the test fails whenever $f(g^n(\mathbf{Z})) = b$ and $f(\mathbf{Z}^j) = b$ for all j, which has probability at least $1-(m+1)\eta \geq 1/2$; this is excluded by $\delta'_g \leq 1/2$. So $g(b,\dots,b) = b$ and the constant b is a polymorphism. The test fails whenever $g^n(\mathbf{Z}) \in A$ while $\mathbf{Z}^j \notin A$ for all j. Since $g^n(\mathbf{Z})$ is uniform, Lemma 8.15 gives

$$\delta \;\geq\; \eta - \sum_{j=1}^{m} \Pr\left[g^n(\mathbf{Z}) \in A,\ \mathbf{Z}^j \in A\right] \;\geq\; \eta - m\,\eta^{2/(1+r)} = \eta\big(1 - m\,\eta^{\frac{1-r}{1+r}}\big) \;\geq\; \frac{\eta}{2},$$

because $\eta \leq \gamma \leq (2m)^{-(1+r)/(1-r)}$. Hence $\Pr[f \neq b] \leq 2\delta$.

Case B: f is γ-far from constants. Fix $v \in B_{j^*}$. Then $P_{j^*} \geq \prod_{k \neq j^*} \Pr[f = v_k] \geq \gamma^{m-1}$, and Lemma 8.9 gives a set T with $|T| \leq L := 2^{m+2}\ln(2\gamma^{-(m-1)})$ and $\beta := \Pr[f \neq \mathrm{rd}_T f] \leq 2^{m+2}\gamma^{-(m-1)}\delta$. Let $F_0 = \cdots = F_m = \mathrm{rd}_T f$ (recall $p_g = 1/2$, so $\mathrm{rd}_T^{p_g} = \mathrm{rd}_T$). If $\delta'_g \leq 2^{-mL}/(1 + (m+1)2^{m+2}\gamma^{-(m-1)})$, then $\delta + (m+1)\beta < 2^{-mL} \leq 2^{-m|T|}$, and Lemma 8.13 shows that $\mathrm{rd}_T f$ is a polymorphism of g at distance $\beta \leq 2^{m+2}\gamma^{-(m-1)}\delta$ from f.

So for non-affine g the theorem holds with $\delta'_g = 2^{-mL}/(1 + (m+1)2^{m+2}\gamma^{-(m-1)})$ and $C'_g = \max\{1/\delta'_g,\, 2^{m+2}\gamma^{-(m-1)}\}$ (Case A needs $C'_g \geq 2$), where m is now the number of relevant coordinates. □

Applied with $\delta = 0$, the proof shows that every polymorphism of a non-affine g with $p_g = 1/2$ that depends on all its coordinates is either constant (Case A) or a junta on at most L coordinates (Case B, where $\beta = 0$), with $L = L(g)$ independent of n; and Lemma 8.14 with $\delta = 0$ shows that every polymorphism of $x_1 \oplus \cdots \oplus x_m \oplus b$, $m \geq 2$, is of the form $\bigoplus_{i \in S} x_i \oplus a$.

Proposition 8.16 (Linear dependence is optimal)**.** *Let $p_g = 1/2$, let g depend on at least two coordinates, and let δ_g be as in Corollary 8.3. Then $\delta_g(\varepsilon) \leq 4(m+1)(1+2C'_g)\,\varepsilon$ for $0 < \varepsilon \leq 1/(8(1+2C'_g))$.*

Proof. For $h\colon \{0,1\}^n \to \{0,1\}$ let $d(h) = \min_F \Pr[h \neq F]$ over the polymorphisms F of g on n variables (uniform measure).

(i) *If f is ε-close to a skew polymorphism (F_0, F_1), then $d(f) \leq (1+2C'_g)\varepsilon$.* As $p_g = 1/2$, both distances are uniform, so $\Pr[F_0 \neq F_1] \leq 2\varepsilon$. Since $F_0(g^n(\mathbf{Z})) = g(F_1(\mathbf{Z}^1),\dots,F_1(\mathbf{Z}^m))$ and $g^n(\mathbf{Z})$ is uniform, $\mathrm{err}_g(F_1,\dots,F_1) \leq \Pr[F_1 \neq F_0] \leq 2\varepsilon$. By Theorem 8.2, F_1 is $2C'_g\varepsilon$-close to a polymorphism, and f is ε-close to F_1.

(ii) $\mathrm{err}_g(f,\dots,f) \leq (m+1)d(f)$. If F is a polymorphism, the test can fail only if $f \neq F$ at one of the $m+1$ uniformly distributed points $g^n(\mathbf{Z}), \mathbf{Z}^1, \dots, \mathbf{Z}^m$.

(iii) *For all large n there is h^* on n variables with $d(h^*) \geq 1/4$.* As in the proof of Theorem 8.2 we may assume that g depends on all $m \geq 2$ coordinates. If g is affine, all its polymorphisms have the form $\bigoplus_{i \in S} x_i \oplus a$; take $h^* = x_1 x_2$. If $S \not\subseteq \{1,2\}$, then $\bigoplus_{i \in S} x_i \oplus a$ is a uniform bit given (x_1, x_2), so it differs from h^* with probability $1/2$; otherwise it is one of the eight affine functions of (x_1, x_2), each of which differs from $x_1 x_2$ on one or three of the four points. If g is not affine, all its polymorphisms are constants or juntas on at most L coordinates; for $n > L$ take $h^* = x_1 \oplus \cdots \oplus x_n$, which is a uniform bit given any L coordinates, so that $d(h^*) = 1/2$.

(iv) Let $\theta = 2(1+2C'_g)\varepsilon \leq 1/4$, and let n be as in (iii) with $2^{-n} \leq \theta$. Change the dictator x_1 into h^* one point at a time. Each step changes d by at most 2^{-n}, $d(x_1) = 0$ and $d(h^*) \geq \theta$, so at the first step where $d \geq \theta$ we obtain f with $\theta \leq d(f) < \theta + 2^{-n} \leq 2\theta$. By (ii), $\mathrm{err}_g(f,\dots,f) \leq 2(m+1)\theta = 4(m+1)(1+2C'_g)\varepsilon$. By (i), f is not ε-close to any skew polymorphism, since $(1+2C'_g)\varepsilon = \theta/2 < d(f)$. Hence $\delta_g(\varepsilon) \leq \mathrm{err}_g(f,\dots,f)$. □

8.6 The exponent for unbalanced g

For unbalanced g the loss is polynomial (Theorem 8.1) but in general not linear. The exponent is governed by the number of coordinates of a constant input that must be flipped to change the value of g. For $b \in \{0,1\}$ write $b^m = (b,\dots,b)$, and for $S \subseteq [m]$ let $b^m \oplus \mathbf{1}_S$ be b^m with the coordinates in S flipped.

Definition 8.17. For non-constant $g\colon \{0,1\}^m \to \{0,1\}$ and $b \in \{0,1\}$ let

$$s_b(g) = \min\{|S| :\ S \subseteq [m],\ g(b^m \oplus \mathbf{1}_S) \neq g(b^m)\}, \qquad e(g) = \max\{s_0(g), s_1(g)\}.$$

We say that g is *completable at* b if $s_b(g) = 1$, or if there are $S \subseteq [m]$ with $|S| = s_b(g)$ and $g(b^m \oplus \mathbf{1}_S) \neq g(b^m)$, an index $j \in S$, and an insensitive context $w \in A_j$ with $w_k = 1 - b$ for all $k \in S \setminus \{j\}$.

For example, $s_0(\mathrm{AND}_m) = m$ and $s_1(\mathrm{AND}_m) = 1$, and AND_m ($m \geq 2$) is not completable at 0: necessarily $S = [m]$, and the only context w with $w_k = 1$ for all $k \neq j$ is sensitive for j. For $g = x_1 \wedge (x_2 \vee x_3)$ we have $s_0 = 2$, $s_1 = 1$, and g is completable at 0 (take $S = \{1,2\}$, $j = 2$ and $w = (w_1, w_3) = (1,1)$). If g depends on coordinates $K \subsetneq [m]$ only and G is the induced function, then $s_b(g) = s_b(G)$, g is completable at b iff G is, and $\delta_g = \delta_G$ (as in the proof of Theorem 8.2, g and G have the same approximate and exact skew polymorphisms).

Theorem 8.18 (The exponent for unbalanced g)**.** *Let $g\colon \{0,1\}^m \to \{0,1\}$ depend on all its coordinates and satisfy $p_g \neq 1/2$ (so $m \geq 2$ and g is not affine), and let δ_g be as in Corollary 8.3.*

(a) For $b \in \{0,1\}$, $s = s_b(g)$ and $0 < \varepsilon \leq 1/16$ we have $\delta_g(\varepsilon) \leq \left(4^s \binom{m}{s} + 1\right)\varepsilon^s$. In particular $\delta_g(\varepsilon) = O_g(\varepsilon^{e(g)})$.

(b) There is a constant C_g such that every δ-approximate polymorphism of g is $C_g\delta^{1/e(g)}$-close to a skew polymorphism of g. Hence $\delta_g(\varepsilon) = \Theta_g(\varepsilon^{e(g)})$.

(c) If $g \in \{\mathrm{AND}_m, \mathrm{OR}_m, \mathrm{NAND}_m, \mathrm{NOR}_m\}$, then $e(g) = m$ and $\varepsilon^m/(c_0(2m)^m) \leq \delta_g(\varepsilon) \leq (4^m + 1)\varepsilon^m$ for $0 < \varepsilon \leq 1/16$.

(d) The proof of (b) gives an explicit C_g if, for each $b \in \{0,1\}$ with $s_b(g) < m$, g is completable at b or $\lambda_S > 0$ for some minimal flip set S at b, where λ_S is defined in Lemma 8.23. This holds for every g with $m \leq 4$. The functions g as above that are neither completable at both b nor of the form in (c) are: none for $m = 2$ (out of 8); four for $m = 3$ (out of 160), namely $\mathbf{1}\{|x| = 1\}$, $\mathbf{1}\{|x| = 2\}$ and their negations; and 44 for $m = 4$ (out of 51,976).

Part (d) is a finite check, done by exact enumeration with two scripts in the directory `numerics` of the accompanying repository, `exponent_classification.py` and `doeblin_check.py`. Completability gives an elementary proof of the near-constant case of (b) with explicit constants (Lemma 8.21). The general case uses a strong data processing inequality (Lemmas 8.22 and 8.23).

Corollary 8.19 (When is the loss linear?)**.** *Let g depend on at least two coordinates. Then $\delta_g(\varepsilon) = \Theta_g(\varepsilon)$ if and only if $p_g = 1/2$ or $s_0(g) = s_1(g) = 1$; otherwise $\delta_g(\varepsilon) = \Theta_g(\varepsilon^{e(g)})$ with $e(g) \geq 2$. In particular, for $m = 2$: $\delta_g(\varepsilon) = \Theta(\varepsilon^2)$ for $g \in \{\mathrm{AND}_2, \mathrm{OR}_2, \mathrm{NAND}_2, \mathrm{NOR}_2\}$, and $\delta_g(\varepsilon) = \Theta_g(\varepsilon)$ for the other six functions that depend on both coordinates ($x \oplus y$, $1 \oplus x \oplus y$, $x \wedge \neg y$, $\neg x \wedge y$, $x \vee \neg y$, $\neg x \vee y$).*

Proof. By the remark after Definition 8.17 we may assume that g depends on all $m \geq 2$ coordinates. If $p_g = 1/2$, use Theorem 8.2 and Proposition 8.16. If $p_g \neq 1/2$, Theorem 8.18(a),(b) gives $\delta_g(\varepsilon) = \Theta_g(\varepsilon^{e(g)})$, and $e(g) = 1$ iff $s_0 = s_1 = 1$. For $m = 2$, the four functions $x \wedge \neg y, \dots, \neg x \vee y$ have $s_0 = s_1 = 1$. □

Lemma 8.20 (Duality)**.** *For $h\colon \{0,1\}^N \to \{0,1\}$ let $h^*(x) = 1 \oplus h(\mathbf{1} \oplus x)$. Then $p_{g^*} = 1 - p_g$; $s_b(g^*) = s_{1-b}(g)$; g^* is completable at b iff g is completable at $1-b$; $\mathrm{err}_{g^*}(f^*, \dots, f^*) = \mathrm{err}_g(f, \dots, f)$; (F_0, F_1) is a skew polymorphism of g iff (F_0^*, F_1^*) is a skew polymorphism of g^*; and $\Pr_{\mu_q}[h_1^* \neq h_2^*] = \Pr_{\mu_{1-q}}[h_1 \neq h_2]$. Consequently $\delta_{g^*} = \delta_g$. Moreover $\mathrm{AND}_m^* = \mathrm{OR}_m$ and $\mathrm{NAND}_m^* = \mathrm{NOR}_m$.*

Proof. Write $\bar{x} = \mathbf{1} \oplus x$, and let $\bar{\mathbf{Z}}$ be the entrywise complement of $\mathbf{Z}$, which is again uniform. Since $g^*(\mathbf{z}_i) = 1 \oplus g(\bar{\mathbf{z}}_i)$, we have $(g^*)^n(\mathbf{Z}) = \mathbf{1} \oplus g^n(\bar{\mathbf{Z}})$, hence $f^*((g^*)^n(\mathbf{Z})) = 1 \oplus f(g^n(\bar{\mathbf{Z}}))$, while $g^*(f^*(\mathbf{Z}^1), \dots, f^*(\mathbf{Z}^m)) = 1 \oplus g(f(\bar{\mathbf{Z}}^1), \dots, f(\bar{\mathbf{Z}}^m))$. So the test for f^* and g^* at $\mathbf{Z}$ is the test for f and g at $\bar{\mathbf{Z}}$; the same computation applies to pairs (F_0, F_1). If $\mathbf{x} \sim \mu_q$ then $\bar{\mathbf{x}} \sim \mu_{1-q}$; this gives the statement on distances, and $p_{g^*} = 1 - p_g$. Next, $g^*(b^m \oplus \mathbf{1}_S) \oplus g^*(b^m) = g((1-b)^m \oplus \mathbf{1}_S) \oplus g((1-b)^m)$, so $s_b(g^*) = s_{1-b}(g)$ with the same minimal sets S; and a context w is insensitive for j with respect to g^* iff $\bar{w}$ is insensitive for j with respect to g, while $w_k = 1 - b$ iff $\bar{w}_k = 1 - (1-b)$. This gives the statement on completability. Since $f \mapsto f^*$ is a bijection that preserves the error and maps ε-close skew polymorphisms of g to ε-close skew polymorphisms of g^* (distances to F_0 are measured under μ_{p_g} and $\mu_{p_{g^*}} = \mu_{1-p_g}$, respectively), $\delta_{g^*} = \delta_g$. The last claim is De Morgan's law. □

Lemma 8.21 (Near-constant functions)**.** *Let g depend on all its $m \geq 2$ coordinates, $p_g \neq 1/2$, $b \in \{0,1\}$, $s = s_b(g)$, and suppose that g is completable at b. Let $f\colon \{0,1\}^n \to \{0,1\}$, $\delta = \mathrm{err}_g(f, \dots, f)$ and $\eta = \Pr[f \neq b] \leq 1/(4m)$. Then*

$$\eta \leq (2^{m+1}\delta)^{1/s} \qquad \text{and} \qquad \Pr_{\mu_{p_g}}[f \neq g(b^m)] \leq \delta + m\eta.$$

In particular f is $(\delta + m(2^{m+1}\delta)^{1/s})$-close to the skew polymorphism $(g(b^m), b)$ (a pair of constants).

Proof. First, $A_k \neq \emptyset$ for every $k \in [m]$: otherwise flipping coordinate k always flips g, so $g(z) = z_k \oplus g(z^{k \to 0})$ and $p_g = 1/2$. Choose S, j and $w \in A_j$ as in Definition 8.17 (if $s = 1$: any $S = \{j\}$ with $g(b^m \oplus \mathbf{1}_{\{j\}}) \neq g(b^m)$ and any $w \in A_j$). Put $U = S \setminus \{j\}$ and $A = \{x : f(x) \neq b\}$, and let $v \in \{0,1\}^{[m]\setminus\{j\}}$ be given by $v_k = 1 - b$ for $k \in U$ and $v_k = b$ otherwise. Then $v^{j \to b} = b^m \oplus \mathbf{1}_U$ and $v^{j \to 1-b} = b^m \oplus \mathbf{1}_S$, and $g(b^m \oplus \mathbf{1}_U) = g(b^m) \neq g(b^m \oplus \mathbf{1}_S)$ by the minimality of s; so $v \in B_j$. Let $E = \{\mathbf{c}_j = v\}$, i.e. $\mathbf{Z}^k \in A$ for $k \in U$ and $\mathbf{Z}^k \notin A$ for $k \in [m] \setminus S$. As the columns are independent, $\Pr[E] = \eta^{s-1}(1-\eta)^{m-s}$.

Fix a row $i \in [n]$, and let $q = \Pr[\mathbf{x}_i = 1 \mid \mathbf{x} \in A]$ and $r = \Pr[\mathbf{x}_i = 1 \mid \mathbf{x} \notin A]$ for uniform $\mathbf{x}$. Since $r = (\frac{1}{2} - \eta q)/(1-\eta)$, both r and $1-r$ are at least $\frac{1}{2} - \eta$. Conditionally on E, the columns $\mathbf{Z}^k$, $k \neq j$, are independent and uniform on A (for $k \in U$) or on $\{0,1\}^n \setminus A$ (for $k \in [m] \setminus S$). Hence the entries of the context $(\mathbf{z}_i)_{-j}$ are conditionally independent, Bernoulli(q) on U and Bernoulli(r) on $[m] \setminus S$. If $s \geq 2$, the context $b^{[m]\setminus\{j\}}$ is insensitive for j (because $g(b^m \oplus \mathbf{1}_{\{j\}}) = g(b^m)$), and so is w. The two contexts take the constant values b and $1-b$ on $U \neq \emptyset$, so their conditional probabilities are at least $(\frac{1}{2} - \eta)^{m-s}$ times q^{s-1} and $(1-q)^{s-1}$ (in some order). Therefore for $s \geq 2$,

$$\alpha := \Pr[i \in \mathbf{R}_j \mid E] \;\geq\; \Big(\frac{1}{2} - \eta\Big)^{m-s}\big(q^{s-1} + (1-q)^{s-1}\big) \;\geq\; \Big(\frac{1}{2} - \eta\Big)^{m-s} 2^{2-s} = 2^{2-m}(1-2\eta)^{m-s}.$$

If $s = 1$ then $U = \emptyset$, $v = b^{[m]\setminus\{j\}}$ and $\Pr[E] = (1-\eta)^{m-1}$; the context $b^{[m]\setminus\{j\}} = v$ is then sensitive for j, so only w is used, and $\alpha \geq \Pr[(\mathbf{z}_i)_{-j} = w \mid E] \geq (\frac{1}{2} - \eta)^{m-1}$. In both cases $\alpha \geq 2^{1-m}(1-2\eta)^m \geq 2^{1-m}(1-2m\eta) \geq 2^{-m}$.

Since $E \subseteq \{\mathbf{c}_j \in B_j\}$, Lemma 8.8 (with $f_0 = \cdots = f_m = f$) gives

$$\delta \;\geq\; \sum_{S' \neq \emptyset} \hat{f}(S')^2 \Pr[E,\ S' \cap \mathbf{R}_j \neq \emptyset] \;\geq\; \sum_{S' \neq \emptyset} \hat{f}(S')^2 \Pr[E]\,\alpha \;\geq\; \eta(1-\eta) \cdot \eta^{s-1}(1-\eta)^{m-s} \cdot 2^{-m},$$

where for $S' \neq \emptyset$ we picked $i \in S'$ and used $\Pr[E,\ S' \cap \mathbf{R}_j \neq \emptyset] \geq \Pr[E] \Pr[i \in \mathbf{R}_j \mid E]$, and then $\sum_{S' \neq \emptyset} \hat{f}(S')^2 = \mathrm{Var}(f) = \eta(1-\eta)$. As $(1-\eta)^{m-s+1} \geq 1 - m\eta \geq \frac{3}{4}$, we get $\eta^s \leq \frac{4}{3} 2^m \delta \leq 2^{m+1}\delta$.

Finally, if $f(\mathbf{Z}^k) = b$ for all k then the right-hand side of the test is $g(b^m)$; so $f(g^n(\mathbf{Z})) \neq g(b^m)$ implies that the test fails or that $f(\mathbf{Z}^k) \neq b$ for some k. As $g^n(\mathbf{Z}) \sim \mu_{p_g}$, this gives $\Pr_{\mu_{p_g}}[f \neq g(b^m)] \leq \delta + m\eta$. The constant pair $(g(b^m), b)$ is clearly a skew polymorphism. □

Without completability the argument of Lemma 8.21 fails: if $\{f \neq b\}$ is concentrated on a subcube, then on the context event almost all rows may be sensitive for j, and the containment inequality gives nothing. Instead we show directly that conditioning s columns on the small set $\{f \neq b\}$ cannot move the law of $g^n(\mathbf{Z})$ far, measured by relative entropy. For distributions ν, π on a finite set with $\nu \ll \pi$ let $D(\nu\|\pi) = \sum_x \nu(x) \ln(\nu(x)/\pi(x))$. We use the chain rule, the convexity of $D(\cdot\|\pi)$, the joint convexity of D, and the data processing inequality $D(\nu K\|\pi K) \leq D(\nu\|\pi)$ for a channel K; see [7, Chapter 2]. The following lemma is a strong data processing inequality in the sense of Ahlswede and Gács [3] (see also [26]); we include a self-contained proof.

Lemma 8.22 (Contraction of relative entropy)**.** *Let $\mathcal{X}, \mathcal{Y}$ be finite sets with $|\mathcal{X}| \geq 2$, let π be a distribution on $\mathcal{X}$ with full support, and let $K = (K(y \mid x))_{x \in \mathcal{X}, y \in \mathcal{Y}}$ be a stochastic matrix. For a distribution ν on $\mathcal{X}^n$ let $\nu K^{\otimes n}$ be the law of $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ when $\mathbf{x} \sim \nu$ and, given $\mathbf{x}$, the $\mathbf{y}_i$ are independent with laws $K(\cdot \mid \mathbf{x}_i)$. Let*

$$\theta = \theta(\pi, K) = \sup\left\{ \frac{D(\nu K\|\pi K)}{D(\nu\|\pi)} :\ \nu \neq \pi \text{ a distribution on } \mathcal{X} \right\} \;\leq\; 1.$$

(a) For every $n \geq 1$ and every distribution ν on $\mathcal{X}^n$, $D(\nu K^{\otimes n}\|(\pi K)^{\otimes n}) \leq \theta\, D(\nu\|\pi^{\otimes n})$.

(b) Let $\mathcal{Y}' = \{y : \pi K(y) > 0\}$. If the bipartite graph on $\mathcal{X} \sqcup \mathcal{Y}'$ with an edge $\{x, y\}$ whenever $K(y \mid x) > 0$ is connected, then $\theta < 1$.

(c) $\theta \leq 1 - \lambda$, where $\lambda = \sum_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} K(y \mid x)$.

Proof. $\theta \leq 1$ is the data processing inequality.

(a) Let $\mathbf{x} \sim \nu$ and let $\mathbf{y}$ be as in the statement. For a random variable $\mathbf{w}$ write $P_{\mathbf{x}_i|\mathbf{w}}$ for the conditional law of $\mathbf{x}_i$ given $\mathbf{w}$, and similarly for $\mathbf{y}_i$; put $\mathbf{y}_{<i} = (\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})$ and $\mathbf{x}_{<i} = (\mathbf{x}_1, \ldots, \mathbf{x}_{i-1})$. Since $(\pi K)^{\otimes n}$ is a product measure, the chain rule gives

$$D(\nu K^{\otimes n}\|(\pi K)^{\otimes n}) = \sum_{i=1}^{n} \mathbb{E}\big[D(P_{\mathbf{y}_i|\mathbf{y}_{<i}}\|\pi K)\big].$$

Given $\mathbf{x}$, the variable $\mathbf{y}_i$ is drawn from $K(\cdot \mid \mathbf{x}_i)$ independently of $(\mathbf{y}_{<i}, \mathbf{x}_{-i})$, so $P_{\mathbf{y}_i|\mathbf{y}_{<i}} = P_{\mathbf{x}_i|\mathbf{y}_{<i}} K$, and by the definition of θ, $D(P_{\mathbf{y}_i|\mathbf{y}_{<i}}\|\pi K) \leq \theta D(P_{\mathbf{x}_i|\mathbf{y}_{<i}}\|\pi)$ (both sides vanish if $P_{\mathbf{x}_i|\mathbf{y}_{<i}} = \pi$). Given $\mathbf{x}_{<i}$, the vector $\mathbf{y}_{<i}$ is independent of $\mathbf{x}_i$, so $P_{\mathbf{x}_i|\mathbf{y}_{<i}} = \mathbb{E}[P_{\mathbf{x}_i|\mathbf{x}_{<i}} \mid \mathbf{y}_{<i}]$, and the convexity of $D(\cdot\|\pi)$ gives $\mathbb{E}D(P_{\mathbf{x}_i|\mathbf{y}_{<i}}\|\pi) \leq \mathbb{E}D(P_{\mathbf{x}_i|\mathbf{x}_{<i}}\|\pi)$. Summing over i and using the chain rule for the product measure $\pi^{\otimes n}$, $\sum_i \mathbb{E}D(P_{\mathbf{x}_i|\mathbf{x}_{<i}}\|\pi) = D(\nu\|\pi^{\otimes n})$.

(b) If $|\mathcal{Y}'| = 1$ then $\nu K = \pi K$ for every ν, and $\theta = 0$; so let $|\mathcal{Y}'| \geq 2$. Let $(\mathbf{x}, \mathbf{y})$ have law $\pi(x)K(y \mid x)$, and let ρ be the maximum of $\mathbb{E}[\varphi(\mathbf{x})\psi(\mathbf{y})]$ over $\varphi\colon \mathcal{X} \to \mathbb{R}$, $\psi\colon \mathcal{Y}' \to \mathbb{R}$ with $\mathbb{E}\varphi =$

$\mathbb{E}\psi = 0$ and $\mathbb{E}\varphi^2 = \mathbb{E}\psi^2 = 1$ (a maximum over a compact set). If $\rho = 1$, then $\mathbb{E}(\varphi(\mathbf{x}) - \psi(\mathbf{y}))^2 = 0$ for a maximizer, so $\varphi(x) = \psi(y)$ for every edge $\{x, y\}$, and φ is constant by connectivity, contradicting $\mathbb{E}\varphi = 0$, $\mathbb{E}\varphi^2 = 1$. So $\rho < 1$. Norms below are in $L^2(\pi)$ and $L^2(\pi K)$.

For $\nu \neq \pi$ let $\varphi = \nu/\pi - 1$ and $\psi(y) = \mathbb{E}[\varphi(\mathbf{x}) \mid \mathbf{y} = y]$ for $y \in \mathcal{Y}'$. Then $\nu K(y) = \sum_x \pi(x)(1 + \varphi(x))K(y \mid x) = \pi K(y)(1 + \psi(y))$, $\mathbb{E}\psi = \mathbb{E}\varphi = 0$, and $\|\psi\|^2 = \mathbb{E}[\varphi(\mathbf{x})\psi(\mathbf{y})] \leq \rho\|\varphi\|\|\psi\|$, so $\|\psi\| \leq \rho\|\varphi\|$. For $|t| \leq \frac{1}{2}$ we have $(1+t)\ln(1+t) = t + \frac{t^2}{2} + R(t)$ with $|R(t)| \leq \sum_{k\geq 3} \frac{|t|^k}{k(k-1)} \leq \frac{|t|^3}{3}$. Since $\mathbb{E}\varphi = \mathbb{E}\psi = 0$ and $\|\psi\|_\infty \leq \|\varphi\|_\infty$, if $\|\varphi\|_\infty \leq \frac{1}{2}$ then

$$D(\nu\|\pi) = \mathbb{E}(1+\varphi)\ln(1+\varphi) = \tfrac{1}{2}\|\varphi\|^2 + \mathbb{E}R(\varphi), \qquad D(\nu K\|\pi K) = \tfrac{1}{2}\|\psi\|^2 + \mathbb{E}R(\psi).$$

As $|R(t)| \leq |t|^3/3$, $\|\psi\|_\infty \leq \|\varphi\|_\infty$ and $\|\psi\| \leq \|\varphi\|$, we have $|\mathbb{E}R(\varphi)| \leq \|\varphi\|_\infty\|\varphi\|^2/3$ and $|\mathbb{E}R(\psi)| \leq \|\psi\|_\infty\|\psi\|^2/3 \leq \|\varphi\|_\infty\|\varphi\|^2/3$. Hence, if $\|\varphi\|_\infty \leq t_0 := \min\{\frac{1}{2}, (1-\rho^2)/12\}$, then $\|\varphi\|_\infty/3 \leq (1-\rho^2)/36$ and

$$\frac{D(\nu K\|\pi K)}{D(\nu\|\pi)} \leq \frac{\rho^2/2 + (1-\rho^2)/36}{1/2 - (1-\rho^2)/36} = \frac{1 + 17\rho^2}{17 + \rho^2} < 1.$$

The set N of distributions ν with $\|\nu/\pi - 1\|_\infty \geq t_0$ is compact and does not contain π. On N the ratio $r(\nu) = D(\nu K\|\pi K)/D(\nu\|\pi)$ is continuous, because both relative entropies are continuous in ν (finite sums of terms $u\ln(u/c)$ with $c > 0$) and $D(\nu\|\pi) > 0$. Moreover $r(\nu) < 1$ for each $\nu \in N$. Indeed, let $\nu_y(x) = \nu(x)K(y \mid x)/\nu K(y)$ and $\pi_y(x) = \pi(x)K(y \mid x)/\pi K(y)$ be the conditional laws of $\mathbf{x}$ given $\mathbf{y} = y$ under the two joint laws $\nu(x)K(y \mid x)$ and $\pi(x)K(y \mid x)$. The chain rule, applied to these joint laws in both orders, gives

$$D(\nu\|\pi) - D(\nu K\|\pi K) = \sum_{y:\, \nu K(y) > 0} \nu K(y)\, D(\nu_y\|\pi_y).$$

If this vanishes, then $1 + \varphi(x) = \nu(x)/\pi(x) = \nu K(y)/\pi K(y) = 1 + \psi(y)$ for every edge $\{x, y\}$ with $\nu K(y) > 0$. If $\nu K(y) = 0$, every neighbour x of y has $\nu(x) = 0$, so again $1 + \varphi(x) = 0 = 1 + \psi(y)$. Thus $\varphi(x) = \psi(y)$ on every edge, φ is constant by connectivity, and $\nu = \pi$, a contradiction. Hence $\theta \leq \max\{(1 + 17\rho^2)/(17 + \rho^2), \max_N r\} < 1$.

(c) We may assume $0 < \lambda < 1$; if $\lambda = 1$ then $K(\cdot \mid x)$ does not depend on x and $\theta = 0$. Let $R(y) = \min_x K(y \mid x)/\lambda$ and $K'(y \mid x) = (K(y \mid x) - \lambda R(y))/(1-\lambda)$; both are stochastic. Then $\nu K = \lambda R + (1-\lambda)\nu K'$ and $\pi K = \lambda R + (1-\lambda)\pi K'$, so joint convexity and data processing give $D(\nu K\|\pi K) \leq (1-\lambda)D(\nu K'\|\pi K') \leq (1-\lambda)D(\nu\|\pi)$. □

Lemma 8.23 (Near-constant functions, general case)**.** *Let g depend on all its $m \geq 2$ coordinates, $p_g \neq 1/2$, $b \in \{0,1\}$, and suppose that $s := s_b(g) < m$. Fix $S \subseteq [m]$ with $|S| = s$ and $g(b^m \oplus \mathbf{1}_S) \neq g(b^m)$. For $x \in \{0,1\}^S$ and $u \in \{0,1\}^{[m]\setminus S}$ write (x, u) for the point of $\{0,1\}^m$ with these restrictions, and let $K_S(y \mid x) = \Pr_{\mathbf{u}}[g(x, \mathbf{u}) = y]$, with $\mathbf{u}$ uniform on $\{0,1\}^{[m]\setminus S}$. Let π be the uniform distribution on $\{0,1\}^S$, $\theta = \theta(\pi, K_S)$, $\lambda_S = \sum_{y\in\{0,1\}} \min_x K_S(y \mid x)$ and $\tau = (1-\theta)/2$. Then $\theta \leq 1 - \lambda_S$ and $\theta < 1$. Moreover, let $f\colon \{0,1\}^n \to \{0,1\}$, $\delta = \mathrm{err}_g(f, \dots, f)$, and suppose that*

$$\eta = \Pr[f \neq b] \;\leq\; \eta_0 := \min\left\{\frac{\tau}{2m},\ \left(4\tbinom{m}{s} + 4\right)^{-\frac{1+\theta}{(1-\theta)s}}\right\}.$$

Then $\eta^s \leq 2\delta/\tau$ and $\Pr_{\mu_{p_g}}[f \neq g(b^m)] \leq (1 + 2\binom{m}{s}/\tau)\delta$. All this remains true if θ is replaced by any $\theta' \in [\theta, 1)$ in the definitions of τ and η_0, for instance by $\theta' = 1 - \lambda_S$ when $\lambda_S > 0$.

Proof. Lemma 8.22(c) gives $\theta \leq 1 - \lambda_S$. Pick $k \in [m] \setminus S$. As g depends on coordinate k, there is $z \in \{0,1\}^m$ with $g(z) \neq g(z \oplus \mathbf{1}_{\{k\}})$, so $K_S(0 \mid z_S) > 0$ and $K_S(1 \mid z_S) > 0$. As every x is adjacent to 0 or to 1, the bipartite graph of Lemma 8.22(b) is connected, and $\theta < 1$.

Let $p = p_g$, $a = g(b^m)$, $A = \{x : f(x) \neq b\}$, $B = \{y : f(y) \neq a\}$, $\nu_B = \mu_p(B)$, $\mathbf{y} = g^n(\mathbf{Z}) \sim \mu_p$ and $\mathbf{S} = \{k \in [m] : \mathbf{Z}^k \in A\}$. The right-hand side of the test is $g(b^m \oplus \mathbf{1}_{\mathbf{S}})$. If $\eta = 0$ then $\mathbf{S} = \emptyset$, the right-hand side is always a, and $\nu_B = \delta$, so both claims hold; let $\eta > 0$. If $|\mathbf{S}| < s$ it equals a, by the minimality of s, so the test fails whenever $\mathbf{y} \in B$ and $|\mathbf{S}| < s$. The columns are independent, so $\Pr[|\mathbf{S}| \geq s] \leq \binom{m}{s}\eta^s$, and

$$\text{(i)} \qquad \delta \;\geq\; \nu_B - \tbinom{m}{s}\eta^s.$$

If $\mathbf{S} = S$, the right-hand side is $1 - a$, so the test fails whenever $\mathbf{S} = S$ and $\mathbf{y} \notin B$. Let E be the event that $\mathbf{Z}^k \in A$ for all $k \in S$, and let P be the conditional law of $\mathbf{y}$ given E. Since $\{\mathbf{S} = S\} \subseteq E$ and $\Pr[\mathbf{S} = S] = \eta^s(1-\eta)^{m-s}$,

$$\text{(ii)} \qquad \delta \;\geq\; \Pr[\mathbf{S} = S] - \Pr[E,\ \mathbf{y} \in B] = \eta^s\big((1-\eta)^{m-s} - P(B)\big).$$

To bound $P(B)$, let $\mathbf{x}_i = (\mathbf{Z}_{ik})_{k \in S} \in \{0,1\}^S$. Unconditionally $(\mathbf{x}_1, \dots, \mathbf{x}_n) \sim \pi^{\otimes n}$, and E is an event of probability η^s determined by $(\mathbf{x}_1, \dots, \mathbf{x}_n)$; let ν be the conditional law of $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ given E, so that $D(\nu \| \pi^{\otimes n}) = \ln(1/\eta^s) = s\ln(1/\eta)$. The entries $\mathbf{Z}_{ik}$, $k \notin S$, are uniform and independent of $(\mathbf{x}_1, \dots, \mathbf{x}_n)$, so given E and the $\mathbf{x}_i$, the bits $\mathbf{y}_i = g(\mathbf{x}_i, (\mathbf{Z}_{ik})_{k \notin S})$ are independent with laws $K_S(\cdot \mid \mathbf{x}_i)$; that is, $P = \nu K_S^{\otimes n}$. Also $(\pi K_S)^{\otimes n} = \mu_p$. By Lemma 8.22(a), $D(P\|\mu_p) \leq \theta s \ln(1/\eta)$. By data processing for the map $y \mapsto \mathbf{1}_B(y)$, and since $u\ln\frac{u}{v} + (1-u)\ln\frac{1-u}{1-v} \geq u\ln\frac{1}{v} - \ln 2$ for $u, v \in [0,1]$,

$$\text{(iii)} \qquad P(B)\ln\frac{1}{\nu_B} \;\leq\; \theta s \ln\frac{1}{\eta} + \ln 2$$

(if $\nu_B = 0$ then $P(B) = 0$, as $P \ll \mu_p$). We now distinguish two cases.

Case 1: $P(B) \leq 1 - \tau$. As $\eta \leq \tau/(2m)$, $(1-\eta)^{m-s} \geq 1 - m\eta \geq 1 - \tau/2$, and (ii) gives $\delta \geq \eta^s\tau/2$.

Case 2: $P(B) > 1 - \tau = (1+\theta)/2$. Then (iii) gives $\ln(1/\nu_B) < \kappa s \ln(1/\eta) + 2\ln 2$ with $\kappa = 2\theta/(1+\theta) < 1$, i.e. $\nu_B > \eta^{\kappa s}/4$. As $\eta \leq \eta_0$ and $1 - \kappa = (1-\theta)/(1+\theta)$, we have $\eta^{(1-\kappa)s} \leq 1/(4\binom{m}{s}+4)$, i.e. $\eta^{\kappa s}/4 \geq (\binom{m}{s}+1)\eta^s$, and (i) gives $\delta \geq \eta^s$.

In both cases $\delta \geq \eta^s\tau/2$, as $\tau \leq 1/2$. Finally (i) gives $\nu_B \leq \delta + \binom{m}{s}\eta^s \leq (1 + 2\binom{m}{s}/\tau)\delta$. The argument used θ only through the bound $D(P\|\mu_p) \leq \theta s \ln(1/\eta)$, which holds for every $\theta' \geq \theta$. $\square$

Proof of Theorem 8.18(b). Functions of AND *type.* Let $g \in \{\mathrm{AND}_m, \mathrm{OR}_m, \mathrm{NAND}_m, \mathrm{NOR}_m\}$, so $e(g) = m$. For every $\varepsilon \in (0, 1/2)$, the proof of (c) (for AND_m, the proof of Corollary 8.3; for NAND_m, the NAND_m argument in the proof of (c)) shows that every $\varepsilon^m/(c_0(2m)^m)$-approximate polymorphism of g is ε-close to a skew polymorphism; for OR_m and NOR_m this follows from the cases AND_m and NAND_m via the bijection $f \mapsto f^*$ of Lemma 8.20, which preserves errors and distances. Given a δ-approximate polymorphism f, put $\varepsilon = 2m(c_0\delta)^{1/m}$. If $\varepsilon < 1/2$ then $\delta = \varepsilon^m/(c_0(2m)^m)$ and f is ε-close to a skew polymorphism; otherwise the dictator pair (x_1, x_1), a skew polymorphism of every g, is at distance at most $1 \leq 2\varepsilon$. So (b) holds with $C_g = 4mc_0^{1/m}$.

Other functions. Now $s_0(g), s_1(g) < m$: $s_0(g) = m$ means that g is constant on $\{0,1\}^m \setminus \{\mathbf{1}\}$ and takes the other value at $\mathbf{1}$, i.e. $g \in \{\mathrm{AND}_m, \mathrm{NAND}_m\}$, and similarly $s_1(g) = m$ forces $g \in \{\mathrm{OR}_m, \mathrm{NOR}_m\}$. For $b \in \{0,1\}$ fix a minimal flip set at b, and let $\eta_0^{(b)}$ and $\tau^{(b)}$ be the constants of Lemma 8.23. Put $\gamma = \min\{1/(4m), \eta_0^{(0)}, \eta_0^{(1)}\}$ and $\tau_* = \min\{\tau^{(0)}, \tau^{(1)}\}$, and let f be a δ-approximate polymorphism, $\delta = \mathrm{err}_g(f, \dots, f)$. If $\Pr[f \neq b] \leq \gamma$ for some b, Lemma 8.23 (at this b) shows that f is $\max\{(2\delta/\tau_*)^{1/s_b}, (1 + 2^{m+1}/\tau_*)\delta\}$-close to the constant skew polymorphism $(g(b^m), b)$. Both terms are at most $(1 + 2^{m+1}/\tau_*)\delta^{1/e(g)}$, because $\delta \leq 1$, $s_b \leq e(g)$ and $2/\tau_* \geq 1$.

Otherwise f is γ-far from constants. Fix $j \in [m]$ and $v \in B_j$. Then $P_j \geq \Pr[\mathbf{c}_j = v] = \prod_{k \neq j}\Pr[f = v_k] \geq \gamma^{m-1}$, and $a_j \geq 2^{1-m}$ (by the first paragraph of the proof of Lemma 8.21). Lemma 8.9 gives $T \subseteq [n]$ with $|T| \leq L := 2^{m+2}\ln(2\gamma^{1-m})$ and $\beta := \Pr[f \neq \mathrm{rd}_T f] \leq 2^{m+2}\gamma^{1-m}\delta$,

and Lemma 8.10 (with $J = \mathrm{rd}_T f$) gives $\beta_0 := \Pr_{\mu_{p_g}}[f \neq \mathrm{rd}_T^{p_g} f] \le 2^m(\delta+\beta)$. Hence $\delta + \beta_0 + m\beta \le K\delta$ with $K = 1 + 2^m + (2^m+m)2^{m+2}\gamma^{1-m}$. If $\delta < 2^{-mL}/K$, Lemma 8.13 shows that $(\mathrm{rd}_T^{p_g} f, \mathrm{rd}_T f, \dots, \mathrm{rd}_T f)$ is a multi-polymorphism, i.e. $(\mathrm{rd}_T^{p_g} f, \mathrm{rd}_T f)$ is a skew polymorphism, and it is $K\delta$-close to f. If $\delta \ge 2^{-mL}/K$, the dictator pair (x_1, x_1) is a skew polymorphism at distance at most $1 \le (K2^{mL}\delta)^{1/e(g)} \le K2^{mL}\delta^{1/e(g)}$. So all cases are covered by $C_g = \max\{1+2^{m+1}/\tau_*,\ K2^{mL}\}$. Thus every δ-approximate polymorphism is $C_g\delta^{1/e(g)}$-close to a skew polymorphism, i.e. $\delta_g(\varepsilon) \ge (\varepsilon/C_g)^{e(g)}$; together with (a), $\delta_g(\varepsilon) = \Theta_g(\varepsilon^{e(g)})$.

Explicit constants (part (d)). The only non-explicit quantity above is $\theta(\pi, K_S)$. If g is completable at b, Lemma 8.21 can replace Lemma 8.23 at b (with $\gamma \le 1/(4m)$), giving distance at most $(1+m2^{m+1})\delta^{1/e(g)}$. If $\lambda_S > 0$, Lemma 8.23 can be applied with θ replaced by $1-\lambda_S$. Part (d) follows, and the census is a finite computation. □

Proof of Theorem 8.18(a). Let $p = p_g$, $s = s_b(g)$ and $a = g(b^m)$. Since g is not affine and depends on all $m \ge 2$ coordinates, the classification of exact skew polymorphisms [6, Theorem 2.4] shows that every skew polymorphism (F_0, F_1) of g on n variables has

$$F_1 \in \mathcal{F}_n := \{0,1\} \cup \{x_i,\ 1 \oplus x_i :\ i \in [n]\} \cup \{\mathrm{AND}_I,\ \mathrm{OR}_I :\ I \subseteq [n]\}.$$

Let $t \ge 1$ be the integer with $2\varepsilon < 2^{-t-1} \le 4\varepsilon$ (it exists as $4\varepsilon \le 1/4$), put $\eta' = 2^{-t-1} \le 1/4$, let $n \ge t+1$, and let $D = \{x \in \{0,1\}^n :\ x_1 = 0,\ x_2 = \cdots = x_{t+1} = 1\}$, so that $\mu_{1/2}(D) = \eta'$.

Claim: $\Pr_{\mu_{1/2}}[\mathbf{1}_D \neq F] \ge \eta'$ *and* $\Pr_{\mu_{1/2}}[1 \oplus \mathbf{1}_D \neq F] \ge \eta'$ *for every* $F \in \mathcal{F}_n$. The second statement is the first one for $1 \oplus F$, and $1 \oplus F$ ranges over $\mathcal{F}_n$ together with the functions $\mathrm{AND}_I(\mathbf{1} \oplus x)$ and $\mathrm{OR}_I(\mathbf{1} \oplus x)$. Writing $\mu(F) = \Pr[F = 1]$ we have $\Pr[\mathbf{1}_D \neq F] = \eta' + \mu(F) - 2\Pr[D \cap \{F = 1\}]$, so it suffices to show $\Pr[D \cap \{F = 1\}] \le \mu(F)/2$ for all these F. This is clear if $\mu(F) \ge 2\eta'$. Otherwise $\mu(F) < 2\eta' \le 1/2$, so $F \equiv 0$, or $F = \mathrm{AND}_I$ or $F = \mathrm{AND}_I(\mathbf{1} \oplus x) = \mathbf{1}\{x_I = 0^I\}$ with $|I| \ge 2$. If $F = \mathrm{AND}_I$ and $1 \in I$ then $D \cap \{F = 1\} = \emptyset$; if $1 \notin I$ then $D \cap \{F = 1\} \subseteq \{x_1 = 0,\ x_I = 1^I\}$, which has measure $\mu(F)/2$. If $F = \mathbf{1}\{x_I = 0^I\}$ and I meets $\{2, \dots, t+1\}$, then $D \cap \{F = 1\} = \emptyset$; otherwise $D \cap \{F = 1\}$ has measure $\eta' 2^{-|I \setminus \{1\}|} \le 2\eta' 2^{-|I|} \le \mu(F)/2$, as $\eta' \le 1/4$.

Let $\theta = (p + \frac{1}{2})/2$, and let $M = \{x : |x| > \theta n\}$ if $p < 1/2$ and $M = \{x : |x| < \theta n\}$ if $p > 1/2$. By Hoeffding's inequality, $\mu_{1/2}(\{0,1\}^n \setminus M) \le \tau_n$ and $\mu_p(M) \le \tau_n$, where $\tau_n = \exp(-(\frac{1}{2} - p)^2 n/2)$. Define

$$f(x) = b \oplus \mathbf{1}_D(x) \quad (x \in M), \qquad f(x) = a \quad (x \notin M).$$

Error. Let G be the event that $g^n(\mathbf{Z}) \notin M$ and $\mathbf{Z}^k \in M$ for all k; as $g^n(\mathbf{Z}) \sim \mu_p$ and $\mathbf{Z}^k \sim \mu_{1/2}$, $\Pr[G] \ge 1 - (m+1)\tau_n$. On G the left-hand side of the test is $a = g(b^m)$ and the right-hand side is $g(b^m \oplus \mathbf{1}_{\mathbf{S}})$ with $\mathbf{S} = \{k : \mathbf{Z}^k \in D\}$; they differ only if $|\mathbf{S}| \ge s$. By independence of the columns, $\Pr[|\mathbf{S}| \ge s] \le \binom{m}{s}\eta'^s$. Hence $\mathrm{err}_g(f, \dots, f) \le \binom{m}{s}(4\varepsilon)^s + (m+1)\tau_n$.

Distance. For every skew polymorphism (F_0, F_1) we have $F_1 \in \mathcal{F}_n$, and by the claim $\Pr_{\mu_{1/2}}[f \neq F_1] \ge \Pr[b \oplus \mathbf{1}_D \neq F_1] - \mu_{1/2}(\{0,1\}^n \setminus M) \ge \eta' - \tau_n$.

Choose n so large that $(m+1)\tau_n \le \varepsilon^s$. Then $\tau_n < \varepsilon < \eta' - \varepsilon$, so f is not ε-close to any skew polymorphism, while $\mathrm{err}_g(f, \dots, f) \le (4^s\binom{m}{s} + 1)\varepsilon^s$. Hence $\delta_g(\varepsilon) \le (4^s\binom{m}{s} + 1)\varepsilon^s$. Applying this with b such that $s_b(g) = e(g)$ gives $\delta_g(\varepsilon) = O_g(\varepsilon^{e(g)})$. □

The construction hides a small set D near the middle layer, where the uniform measure lives, while near the $p_g n$-th layer, where μ_{p_g} lives, f is the constant $g(b^m)$; the test notices D only when at least $s_b(g)$ of the m columns fall into D. This is the mechanism behind the skew polymorphisms of [6], now used quantitatively. The lower bound for the AND-type functions uses a skew version of Theorem 1.4(b).

Lemma 8.24 (Skew m-ary AND test)**.** *Let $m \geq 2$, $F_0, f\colon \{0,1\}^n \to \{0,1\}$, $\mathbf{x}_1, \dots, \mathbf{x}_m$ independent and uniform, $\mu = \mathbb{E}f$ and $\varepsilon' = \Pr[F_0(\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_m) \neq \prod_j f(\mathbf{x}_j)]$. Then $\operatorname{dist}(f) \leq \min\{\mu, c_0\varepsilon'/\mu^{m-1}\} \leq (c_0\varepsilon')^{1/m}$.*

Proof. In the proof of Corollary 3.3, the two m-tuples $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m)$ and $(\mathbf{x}'_1, \mathbf{x}_2, \dots, \mathbf{x}_m)$ have the same meet, hence the same left-hand side $F_0(\cdot)$ whatever F_0 is. So if $f(\mathbf{x}_2) = \cdots = f(\mathbf{x}_m) = 1$ and $f(\mathbf{x}_1) \neq f(\mathbf{x}'_1)$, the right-hand sides differ and one of the two tests fails; this is the only property of the test used there, and the rest of that proof only involves f. So the proof of Corollary 3.3, with the outer function F_0 in place of f, gives $\eta_{f|f} \leq \varepsilon'/\mu^m$ verbatim, and the proof of Theorem 1.4(b) (which uses only this bound, f and Lemma 5.2) applies verbatim and gives $\operatorname{dist}(f) \leq \min\{\mu, c_0\varepsilon'/\mu^{m-1}\}$. □

Proof of Theorem 8.18(c). We have $s_0(\mathrm{AND}_m) = m$, $s_1(\mathrm{AND}_m) = 1$, $s_0(\mathrm{NAND}_m) = m$ and $s_1(\mathrm{NAND}_m) = 1$, so $e = m$ for these two functions, and by Lemma 8.20 also for $\mathrm{OR}_m = \mathrm{AND}^*_m$ and $\mathrm{NOR}_m = \mathrm{NAND}^*_m$. The upper bound is part (a) with $s = m$. By Lemma 8.20 it suffices to prove the lower bound for AND_m and NAND_m; the case of AND_m is Corollary 8.3. Let $g = \mathrm{NAND}_m$, so $p_g = 1 - 2^{-m}$ and $g^n(\mathbf{Z}) = \mathbf{1} \oplus (\mathbf{Z}^1 \wedge \cdots \wedge \mathbf{Z}^m)$. Let f satisfy $\mathrm{err}_g(f, \dots, f) \leq \delta := \varepsilon^m/(c_0(2m)^m)$, and put $\tilde{f}(y) = 1 \oplus f(\mathbf{1} \oplus y)$. The test reads $f(\mathbf{1} \oplus (\mathbf{Z}^1 \wedge \cdots \wedge \mathbf{Z}^m)) = 1 \oplus \prod_k f(\mathbf{Z}^k)$, i.e. $\tilde{f}(\mathbf{Z}^1 \wedge \cdots \wedge \mathbf{Z}^m) = \prod_k f(\mathbf{Z}^k)$. By Lemma 8.24 (with $F_0 = \tilde{f}$) there is $h \in \mathcal{A}_n$ with $\Pr[f \neq h] \leq (c_0\delta)^{1/m} = \varepsilon/(2m)$. As $h(\mathbf{Z}^1 \wedge \cdots \wedge \mathbf{Z}^m) = \prod_k h(\mathbf{Z}^k)$, we get $\Pr_{\mu_{2^{-m}}}[\tilde{f} \neq h] \leq \delta + m\varepsilon/(2m) \leq \varepsilon$, i.e. $\Pr_{\mu_{p_g}}[f \neq \tilde{h}] \leq \varepsilon$ with $\tilde{h}(y) = 1 \oplus h(\mathbf{1} \oplus y)$. If $h = \mathrm{AND}_I$ then $\tilde{h} = \mathrm{OR}_I$, and if $h \equiv 0$ then $\tilde{h} \equiv 1$. In both cases $(\tilde{h}, h)$ is a skew polymorphism of NAND_m: indeed $\mathrm{OR}_I(g^n(\mathbf{Z})) = 1$ iff $\mathbf{Z}_{ik} = 0$ for some $i \in I$ and $k \in [m]$, iff $\mathrm{AND}_I(\mathbf{Z}^k) = 0$ for some k, iff $g(\mathrm{AND}_I(\mathbf{Z}^1), \dots, \mathrm{AND}_I(\mathbf{Z}^m)) = 1$; and $(1, 0)$ is a constant skew polymorphism since $g(0^m) = 1$. So f is ε-close to a skew polymorphism, and $\delta_{\mathrm{NAND}_m}(\varepsilon) \geq \varepsilon^m/(c_0(2m)^m)$. □

Remark 8.25. The exponent c_g is far from optimal; for $g = \mathrm{AND}_2$ ($k = 2$) it gives $c_g = 1/72$, while the truth is $1/2$ (Theorems 5.1 and 1.2). The only non-elementary input is the Chernoff bound: Jones' regularity lemma and the "It Ain't Over Till It's Over" theorem used in [6, 2] are replaced by the containment inequality of Lemma 8.8, which shows that f_{j^*} is close to a junta on $O(\log(1/\varepsilon))$ coordinates with polynomial error. With this junta at hand, the remaining steps follow the outline of [6, Section 4] with explicit polynomial losses. Corollary 8.3 shows that the conjecture of [2] holds in its "polynomial" form for every functional predicate under the uniform distribution, while its "even linear" form fails already (an observation that is known in substance, cf. [16, 17, 6]; we claim no novelty for it) for the predicate $\{(a, b, a \wedge b)\}$ with the uniform distribution on (a, b). Indeed, if $g_3(x \wedge y) = g_1(x)g_2(y)$ for all x, y and $g_1, g_2 \not\equiv 0$, then putting $y = \mathbf{1}$ (resp. $x = \mathbf{1}$) gives $g_3 = g_2(\mathbf{1})g_1$ and $g_3 = g_1(\mathbf{1})g_2$, whence $g_1(\mathbf{1}) = g_2(\mathbf{1}) = 1$ (otherwise $g_3 \equiv 0$ and $g_1 g_2 \equiv 0$) and $g_1 = g_2 = g_3 \in \mathcal{A}_n$. So every exact solution (g_1, g_2, g_3) has $g_1 \in \mathcal{A}_n$ or $g_2 \in \mathcal{A}_n$. Hence, by Theorem 1.2, for every $\delta \leq 1/4$ and $\eta > 0$ there is f such that $f_1 = f_2 = f_3 = f$ has error $\varepsilon(f) \leq \delta$ and is at distance at least $\operatorname{dist}(f) \geq (1 - \eta)\sqrt{\delta}$ from every exact solution.

8.7 Multi-polymorphisms: the optimal exponent

Theorem 8.1 also shows that the dependence in [6, Theorem 4.1], for multi-polymorphisms, is polynomial. We now determine its optimal exponent. A tuple $\mathbf{f}$ is *ε-close* to a tuple $\mathbf{F}$ if $\Pr_{\mu_{p_g}}[f_0 \neq F_0] \leq \varepsilon$ and $\Pr[f_j \neq F_j] \leq \varepsilon$ for all $j \in [m]$. Let $\delta_g^{\mathrm{multi}}(\varepsilon)$ be the supremum of the δ such that, for every n, every δ-approximate multi-polymorphism of g is ε-close to a multi-polymorphism of g. The columns of a tuple need not be equal, so some of them can be close to constants while the

others are not. As a result the exponent is governed by all restrictions of g, and not only by the two constant inputs. For $z \in \{0,1\}^m$ and $S \subseteq [m]$ let $z \oplus \mathbf{1}_S$ be z with the coordinates in S flipped.

Definition 8.26. Let $g\colon \{0,1\}^m \to \{0,1\}$ be non-constant. A *flip pair* of g is a pair (z, S) with $z \in \{0,1\}^m$ and $S \subseteq [m]$ such that $g(z \oplus \mathbf{1}_S) \neq g(z)$ and $g(z \oplus \mathbf{1}_{S'}) = g(z)$ for every $S' \subsetneq S$. Let $E(g) = \max\{|S| :\ (z, S) \text{ is a flip pair of } g\}$.

So (z, S) is a flip pair iff the restriction of g to the subcube $\{z \oplus \mathbf{1}_{S'} : S' \subseteq S\}$ is $g(z) \oplus \prod_{k \in S}(x_k \oplus z_k)$, an AND of $|S|$ literals or the negation of one. Thus $E(g)$ is the largest k such that fixing $m-k$ of the variables of g can leave an AND of k literals or its negation. We say that g is *of AND type* if $E(g) = m$, that is, if $g(x) = c \oplus \prod_{k=1}^{m}(x_k \oplus a_k)$ for some $c \in \{0,1\}$ and $a \in \{0,1\}^m$. For $m = 2$ these are the eight functions AND_2, OR_2, NAND_2, NOR_2, $x \wedge \neg y$, $\neg x \wedge y$, $x \vee \neg y$ and $\neg x \vee y$.

Lemma 8.27 (Properties of E)**.** *Let $g\colon \{0,1\}^m \to \{0,1\}$ be non-constant.*

(a) If $z \in \{0,1\}^m$, $P \subseteq [m]$, and g is not constant on $\{z \oplus \mathbf{1}_{S'} : S' \subseteq P\}$, then g has a flip pair (z, S) with $\emptyset \neq S \subseteq P$.

(b) $1 \leq E(g) \leq m$, and $E(g) \geq e(g)$.

(c) $E(g) = 1$ if and only if g is affine.

(d) If g depends only on the coordinates in $K \subseteq [m]$ and $G\colon \{0,1\}^K \to \{0,1\}$ is the induced function, then $E(g) = E(G)$.

For example, $E(\mathrm{AND}_m) = m$ and $E(\mathrm{Maj}_3) = 2$; and $E(x \wedge \neg y) = 2$, while $e(x \wedge \neg y) = 1$.

Proof. (a) Choose $S \subseteq P$ minimal under inclusion with $g(z \oplus \mathbf{1}_S) \neq g(z)$. Then $S \neq \emptyset$ and (z, S) is a flip pair.

(b) By (a) with $P = [m]$, $E(g) \geq 1$; clearly $E(g) \leq m$. If $b \in \{0,1\}$, $|S| = s_b(g)$ and $g(b^m \oplus \mathbf{1}_S) \neq g(b^m)$, then (b^m, S) is a flip pair by the minimality of $s_b(g)$. So $E(g) \geq e(g)$.

(c) Let $g(x) = \lambda_0 \oplus \bigoplus_k \lambda_k x_k$, and let (z, S) be a flip pair with $|S| \geq 2$. Then $g(z \oplus \mathbf{1}_{\{k\}}) = g(z)$ gives $\lambda_k = 0$ for every $k \in S$, so $g(z \oplus \mathbf{1}_S) = g(z)$, a contradiction. Hence $E(g) = 1$. Conversely, let $E(g) = 1$, fix z and $j \neq k$, and consider the square $Q = \{z, z \oplus \mathbf{1}_{\{j\}}, z \oplus \mathbf{1}_{\{k\}}, z \oplus \mathbf{1}_{\{j,k\}}\}$. If the four values of g on Q had odd sum, exactly one of them would differ from the other three, and with z' the point of Q opposite to that one, $(z', \{j,k\})$ would be a flip pair. So the sum is even, i.e. $\partial_j g(z) := g(z) \oplus g(z \oplus \mathbf{1}_{\{j\}})$ satisfies $\partial_j g(z) = \partial_j g(z \oplus \mathbf{1}_{\{k\}})$ for all $k \neq j$; also $\partial_j g(z) = \partial_j g(z \oplus \mathbf{1}_{\{j\}})$ by definition. Hence $\partial_j g$ is a constant λ_j, and flipping the coordinates of x one at a time, starting from 0, gives $g(x) = g(0) \oplus \bigoplus_j \lambda_j x_j$.

(d) If (z, S) is a flip pair and $k \in S$, then $g(z \oplus \mathbf{1}_{S \setminus \{k\}}) = g(z) \neq g(z \oplus \mathbf{1}_S)$, so g depends on coordinate k. Hence $S \subseteq K$, and (z_K, S) is a flip pair of G. Conversely, if (z', S) is a flip pair of G, then (z, S) is a flip pair of g for every z with $z_K = z'$.

Finally, $(0^m, [m])$ is a flip pair of AND_m, and $(000, \{1,2\})$ is a flip pair of Maj_3. Maj_3 is not of AND type, since it takes the value 1 four times while an AND of three literals or its negation takes it once or seven times. The function $x \wedge \neg y$ is of AND type, and $s_0 = s_1 = 1$ for it. □

Theorem 8.28 (The optimal exponent for multi-polymorphisms)**.** *Let $g\colon \{0,1\}^m \to \{0,1\}$ depend on all its $m \geq 2$ coordinates.*

(a) $\delta_g^{\mathrm{multi}}(\varepsilon) \leq (2\varepsilon)^{E(g)}$ for $0 < \varepsilon < 2^{-m}$.

(b) There is a constant C_g such that every δ-approximate multi-polymorphism of g is $C_g\delta^{1/E(g)}$-close to a multi-polymorphism of g. Hence $\delta_g^{\mathrm{multi}}(\varepsilon) = \Theta_g(\varepsilon^{E(g)})$.

(c) If g is of AND type, (b) holds with $C_g = 2m+11$; thus $\varepsilon^m/(2m+11)^m \le \delta_g^{\mathrm{multi}}(\varepsilon) \le (2\varepsilon)^m$ for $0 < \varepsilon < 2^{-m}$.

Corollary 8.29 (When is the loss for multi-polymorphisms linear?)**.** *Let g depend on at least two coordinates. Then $\delta_g^{\mathrm{multi}}(\varepsilon) = \Theta_g(\varepsilon)$ if and only if g is affine; otherwise $\delta_g^{\mathrm{multi}}(\varepsilon) = \Theta_g(\varepsilon^{E(g)})$ with $E(g) \ge 2$. For $m = 2$: $\delta_g^{\mathrm{multi}}(\varepsilon) = \Theta(\varepsilon)$ for $x \oplus y$ and $1 \oplus x \oplus y$, and $\Theta(\varepsilon^2)$ for the other eight functions that depend on both coordinates. For Maj_3 and for $x \wedge \neg y$ the loss is linear for polymorphisms (Corollary 8.19) but quadratic for multi-polymorphisms.*

Proof. If g depends only on the coordinates in K and G is the induced function, then $p_G = p_g$, $\mathrm{err}_g(\mathbf{f}) = \mathrm{err}_G(f_0, (f_j)_{j\in K})$, and $\mathbf{F}$ is a multi-polymorphism of g iff $(F_0, (F_j)_{j\in K})$ is one of G, whatever F_j, $j \notin K$, are (as $g^n(\mathbf{Z}) = G^n((\mathbf{Z}^j)_{j\in K})$). Keeping $F_j = f_j$ for $j \notin K$ shows $\delta_g^{\mathrm{multi}} = \delta_G^{\mathrm{multi}}$, and $E(g) = E(G)$ by Lemma 8.27(d). So we may assume that g depends on all $m \ge 2$ coordinates, and the first claim follows from Theorem 8.28 and Lemma 8.27(c). The functions of two variables that depend on both are the two affine ones and the eight of AND type. Finally Maj_3 is balanced and $s_0(x \wedge \neg y) = s_1(x \wedge \neg y) = 1$, while $E = 2$ for both (Lemma 8.27). □

The upper bound is the hiding construction of Theorem 8.18(a); now the constant parts can be put into f_0 and into the columns outside the flip pair. For the lower bound we follow the proof of Theorem 8.18(b), column by column. Columns far from constants are handled by the junta lemma, and near-constant columns by a certificate. In the certificate step the far columns are conditioned on events of constant probability, so only part of the input of the relevant channel is fresh, and we need the following variant of Lemma 8.22. For $u \in [0,1]$ and $v \in (0,1)$ write $d(u\|v) = u \ln \frac{u}{v} + (1-u)\ln\frac{1-u}{1-v}$ for the relative entropy between Bernoulli distributions.

Lemma 8.30 (Partial strong data processing)**.** *Let $\mathcal{X}, \mathcal{U}$ be finite sets with $|\mathcal{X}| \ge 2$, let π_X, π_U be distributions on them with full support, let $h\colon \mathcal{X}\times\mathcal{U} \to \{0,1\}$, and let $K(y \mid x) = \Pr_{\mathbf{u}\sim\pi_U}[h(x,\mathbf{u}) = y]$. Let $p = (\pi_X K)(1)$ and $\theta = \theta(\pi_X, K)$ (Lemma 8.22); assume $0 < p < 1$ and $\theta < 1$, and put*

$$M = \frac{9}{p^2(1-p)^2}, \qquad \theta_1 = \frac{1+\theta}{2}, \qquad C_1 = 1 + \max\Big\{2,\ \frac{M}{1-\theta}\Big\}.$$

Let ρ_X and ρ_U be distributions on $\mathcal{X}^n$ and $\mathcal{U}^n$, let $(\mathbf{x},\mathbf{u}) \sim \rho_X \otimes \rho_U$, and let Q be the law of $\mathbf{y} = (h(\mathbf{x}_1,\mathbf{u}_1), \dots, h(\mathbf{x}_n,\mathbf{u}_n)) \in \{0,1\}^n$. Then

$$D(Q\|\mu_p) \ \le\ \theta_1\, D(\rho_X\|\pi_X^{\otimes n}) + C_1\, D(\rho_U\|\pi_U^{\otimes n}).$$

This remains true if θ is replaced by any $\theta' \in [\theta, 1)$ in the definitions of θ_1 and C_1.

If $\rho_U = \pi_U^{\otimes n}$, then $Q = \rho_X K^{\otimes n}$ and Lemma 8.22(a) gives the factor θ; the point of Lemma 8.30 is that a perturbation of the fresh part costs only a constant factor.

Proof. Step 1: three binary inequalities. Let $r \in (0,1)$, $q \in [0,1]$ and $\ell = \ln\frac{r(1-p)}{p(1-r)}$. Expanding the definitions,

$$d(q\|p) = d(r\|p) + d(q\|r) + \ell\,(q-r). \tag{4}$$

For $s \in \mathbb{R}$ let $r_s = re^s/(1-r+re^s)$. Expanding again, $d(q\|r) - d(q\|r_s) = sq - \ln(1-r+re^s)$, and therefore

$$s(q-r) \ \le\ d(q\|r) + \Lambda(s), \qquad \Lambda(s) = \ln(1-r+re^s) - rs. \tag{5}$$

Here $\Lambda(0) = \Lambda'(0) = 0$ and $\Lambda''(t) = r_t(1-r_t) = r(1-r)e^t/(1-r+re^t)^2 \le r(1-r)e^{|t|}$, because $1-r+re^t \ge \min\{1, e^t\}$. Hence $\Lambda(s) \le r(1-r)(e^{|s|} - 1 - |s|) \le \frac{1}{2}r(1-r)s^2e^{|s|}$. The third inequality is

$$r(1-r)\,\ell^2 e^{|\ell|/2} \ \le\ M\, d(r\|p). \tag{6}$$

Replacing (p, r) by $(1-p, 1-r)$ changes the sign of ℓ and leaves both sides unchanged, so we may assume $L := \ell \geq 0$. Let $\varphi(t) = \ln(1-p+pe^t)$ and $p_t = \varphi'(t) = pe^t/(1-p+pe^t)$, so that $p_0 = p$, $p_L = r$ and $\varphi''(t) = p_t(1-p_t)$. The identity before (5), with (r, q, s) replaced by (p, r, L), gives $d(r\|p) = Lr - \varphi(L) = \int_0^L t\, p_t(1-p_t)\, dt$ (integrate by parts). Write $\lambda(v) = \ln \frac{v}{1-v}$, so that $\lambda(p_t) = \lambda(p) + t$. As a function of $x = \lambda(v)$ we have $v(1-v) = e^x/(1+e^x)^2$, whose logarithmic derivative $1-2v$ lies in $(-1, 1)$, and $v(1-v) \leq e^{-|x|}$. Hence $p_t(1-p_t) \geq p_{t'}(1-p_{t'})e^{-|t-t'|}$. If $L \leq 1$, this gives $d(r\|p) \geq r(1-r)e^{-1}\int_0^L t\, dt = r(1-r)L^2/(2e)$, so $r(1-r)L^2e^{L/2} \leq 2e^{3/2}d(r\|p) \leq 9\, d(r\|p)$. If $L > 1$, then $d(r\|p) \geq p(1-p)\int_0^1 te^{-t}dt = (1-2/e)p(1-p) \geq p(1-p)/4$; moreover $r(1-r) \leq e^{-|\lambda(p)+L|} \leq e^{|\lambda(p)|}e^{-L} \leq e^{-L}/(p(1-p))$, because $e^{|\lambda(p)|} = \max(p, 1-p)/\min(p, 1-p)$. So $r(1-r)L^2e^{L/2} \leq L^2e^{-L/2}/(p(1-p)) \leq 16e^{-2}/(p(1-p)) \leq 9\, d(r\|p)/(p(1-p))^2$. In both cases (6) holds.

Step 2: one coordinate. Let σ be a distribution on $\mathcal{X} \times \mathcal{U}$ with $\mathcal{X}$-marginal σ_X, and let $q = \sigma(h = 1)$. We show

$$d(q\|p) \;\leq\; \theta_1 D(\sigma_X\|\pi_X) + C_1 D(\sigma\|\sigma_X \otimes \pi_U). \tag{7}$$

Let $r = (\sigma_X K)(1)$, the probability that $h = 1$ under $\sigma_X \otimes \pi_U$. By the definition of θ, $d(r\|p) \leq \theta D(\sigma_X\|\pi_X)$. By data processing for the map h, $d(q\|r) \leq D(\sigma\|\sigma_X \otimes \pi_U)$, which is finite because $\sigma \ll \sigma_X \otimes \pi_U$. If $r \in \{0, 1\}$, finiteness forces $q = r$, and (7) follows from $d(q\|p) = d(r\|p)$. Otherwise let $C' = C_1 - 1 = \max\{2, M/(1-\theta)\}$. By (5) with $s = \ell/C'$, then $C' \geq 2$ and (6),

$$\ell(q-r) \leq C'd(q\|r) + C'\Lambda(\ell/C') \leq C'd(q\|r) + \frac{r(1-r)\ell^2 e^{|\ell|/2}}{2C'} \leq C'd(q\|r) + \frac{M}{2C'}\, d(r\|p).$$

With (4) and $M/(2C') \leq (1-\theta)/2$ this gives $d(q\|p) \leq (1 + \frac{1-\theta}{2})\theta D(\sigma_X\|\pi_X) + C_1 D(\sigma\|\sigma_X \otimes \pi_U)$, and $(1 + \frac{1-\theta}{2})\theta \leq \theta + \frac{1-\theta}{2} = \theta_1$.

Step 3: tensorization. For random variables $\mathbf{a}, \mathbf{w}$ write $P_{\mathbf{a}|\mathbf{w}}$ for the conditional law of $\mathbf{a}$ given $\mathbf{w}$, and put $\mathbf{y}_{<i} = (\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})$, and similarly $\mathbf{x}_{<i}$, $\mathbf{u}_{<i}$. By the chain rule for the product measure μ_p, $D(Q\|\mu_p) = \sum_i \mathbb{E}\, d(Q(\mathbf{y}_i = 1 \mid \mathbf{y}_{<i})\|p)$. Apply (7) to $\sigma = P_{(\mathbf{x}_i, \mathbf{u}_i)|\mathbf{y}_{<i}}$, and use $D(\sigma\|\sigma_X \otimes \pi_U) = \mathbb{E}_{x\sim\sigma_X} D(\sigma(\cdot \mid x)\|\pi_U)$:

$$\mathbb{E}\, d(Q(\mathbf{y}_i = 1 \mid \mathbf{y}_{<i})\|p) \leq \theta_1 \mathbb{E} D(P_{\mathbf{x}_i|\mathbf{y}_{<i}}\|\pi_X) + C_1 \mathbb{E} D(P_{\mathbf{u}_i|\mathbf{x}_i, \mathbf{y}_{<i}}\|\pi_U).$$

The vector $\mathbf{y}_{<i}$ is a function of $(\mathbf{x}_{<i}, \mathbf{u}_{<i})$, and the blocks $\mathbf{x}$ and $\mathbf{u}$ are independent, so $P_{\mathbf{x}_i|\mathbf{x}_{<i}, \mathbf{u}_{<i}} = P_{\mathbf{x}_i|\mathbf{x}_{<i}}$ and $P_{\mathbf{u}_i|\mathbf{x}_{\leq i}, \mathbf{u}_{<i}} = P_{\mathbf{u}_i|\mathbf{u}_{<i}}$. By the tower property, $P_{\mathbf{x}_i|\mathbf{y}_{<i}} = \mathbb{E}[P_{\mathbf{x}_i|\mathbf{x}_{<i}} \mid \mathbf{y}_{<i}]$ and $P_{\mathbf{u}_i|\mathbf{x}_i, \mathbf{y}_{<i}} = \mathbb{E}[P_{\mathbf{u}_i|\mathbf{u}_{<i}} \mid \mathbf{x}_i, \mathbf{y}_{<i}]$. By the convexity of $D(\cdot\|\pi)$, the two expectations on the right are at most $\mathbb{E} D(P_{\mathbf{x}_i|\mathbf{x}_{<i}}\|\pi_X)$ and $\mathbb{E} D(P_{\mathbf{u}_i|\mathbf{u}_{<i}}\|\pi_U)$. Summing over i and using the chain rule for $\pi_X^{\otimes n}$ and $\pi_U^{\otimes n}$ proves the lemma. The proof used θ only through $d(r\|p) \leq \theta D(\sigma_X\|\pi_X)$ and $\theta \leq 1$, so any $\theta' \in [\theta, 1)$ can replace it. The constant M in (6) is far from optimal; the applications below only need some finite M that depends on p alone. □

Lemma 8.31 (Approximate multi-polymorphisms of AND_m)**.** *Let $m \geq 2$, let $h_0, \ldots, h_m$ be functions from $\{0,1\}^n$ to $\{0,1\}$, let $\mathbf{W}^1, \ldots, \mathbf{W}^m$ be independent and uniform on $\{0,1\}^n$, and let $\delta = \Pr[h_0(\mathbf{W}^1 \wedge \cdots \wedge \mathbf{W}^m) \neq \prod_{k=1}^m h_k(\mathbf{W}^k)]$. Then there are functions $H_0, \ldots, H_m$ with $H_0(w^1 \wedge \cdots \wedge w^m) = \prod_k H_k(w^k)$ for all $w^1, \ldots, w^m \in \{0,1\}^n$, such that $\Pr_{\mu_{2^{-m}}}[h_0 \neq H_0] \leq (2m+11)\delta^{1/m}$ and $\Pr[h_k \neq H_k] \leq (2m+11)\delta^{1/m}$ for every $k \in [m]$.*

Proof. Let $\mu_k = \mathbb{E} h_k$ and $t = 10\, \delta^{1/m}$, and recall that $\mathbf{W}^1 \wedge \cdots \wedge \mathbf{W}^m \sim \mu_{2^{-m}}$.

If $\mu_{k_0} \leq t$ for some $k_0 \in [m]$, let $H_0 \equiv 0$, $H_{k_0} \equiv 0$ and $H_k = h_k$ for $k \neq k_0$. This tuple is exact, $\Pr[h_{k_0} \neq H_{k_0}] = \mu_{k_0} \leq t$, and $\Pr_{\mu_{2^{-m}}}[h_0 \neq 0] \leq \delta + \Pr[\prod_k h_k(\mathbf{W}^k) = 1] \leq \delta + t$.

Otherwise $\mu_k > t$ for all k. Fix $k \neq l$ in $[m]$. For $w = (w^j)_{j\notin\{k,l\}}$ let $v_w = \bigwedge_{j\notin\{k,l\}} w^j$ (so $v_w = \mathbf{1}$ if $m = 2$) and $\varepsilon_w = \Pr[h_0(\mathbf{W}^k \wedge \mathbf{W}^l \wedge v_w) \neq h_k(\mathbf{W}^k)h_l(\mathbf{W}^l)]$. If $\prod_{j\notin\{k,l\}} h_j(w^j) = 1$, then ε_w is the conditional failure probability of the test given $\mathbf{W}^j = w^j$ for $j \notin \{k,l\}$. As $\Pr[\prod_{j\notin\{k,l\}} h_j(\mathbf{W}^j) = 1] = \prod_{j\notin\{k,l\}} \mu_j \geq t^{m-2}$, there is such a w with $\varepsilon_w \leq \delta/t^{m-2}$, and $\sqrt{\delta/t^{m-2}} = 10^{-(m-2)/2}\delta^{1/m}$. Apply Theorem 1.4(c) with $p = \rho = \frac{1}{2}$ (so $K = 10$) to $F = h_0(\cdot \wedge v_w)$, $G = h_k$ and $H = h_l$. Case (i) would give $\min(\mu_k, \mu_l) \leq 10\sqrt{\varepsilon_w} \leq t$, which is excluded. So case (ii) gives $T_{kl} \subseteq [n]$ with $\Pr[h_k \neq \mathrm{AND}_{T_{kl}}] \leq r$ and $\Pr[h_l \neq \mathrm{AND}_{T_{kl}}] \leq r$, where $r = \frac{6}{5}\sqrt{\delta/t^{m-2}} \leq \frac{6}{5}\delta^{1/m}$; in particular $t \geq 5r$.

All the sets T_{kl} coincide. Indeed, suppose that h_k is r-close to AND_T and to $\mathrm{AND}_{T'}$ with $T \neq T'$, say $|T| \leq |T'|$. Then $T' \not\subseteq T$, and $\Pr[\mathrm{AND}_T \neq \mathrm{AND}_{T'}] \geq 2^{-|T|} - 2^{-|T\cup T'|} \geq 2^{-|T|-1} \geq (\mu_k - r)/2 > (t-r)/2 \geq 2r$, since $2^{-|T|} = \mathbb{E}\mathrm{AND}_T \geq \mu_k - r$. This contradicts $\Pr[\mathrm{AND}_T \neq \mathrm{AND}_{T'}] \leq 2r$. So $T_{kl} = T_{kl'}$ whenever k, l, l' are distinct, and as $T_{kl} = T_{lk}$, all T_{kl} equal one set T. Let $H_0 = \cdots = H_m = \mathrm{AND}_T$; this tuple is exact, $\Pr[h_k \neq \mathrm{AND}_T] \leq r$, and $\Pr_{\mu_{2^{-m}}}[h_0 \neq \mathrm{AND}_T] \leq \delta + \sum_k \Pr[h_k(\mathbf{W}^k) \neq \mathrm{AND}_T(\mathbf{W}^k)] \leq \delta + mr$.

In both cases all distances are at most $\delta + \max\{t, mr\} \leq (1 + 10 + \frac{6m}{5})\delta^{1/m} \leq (2m+11)\delta^{1/m}$. □

Proof of Theorem 8.28(a). Let $s = E(g)$, let (z, S) be a flip pair with $|S| = s$, and let $c = g(z)$ and $p = p_g$. Let $\eta = 2^{-t-1}$ be the power of two in $(\varepsilon, 2\varepsilon]$; as $2\varepsilon < 2^{1-m} \leq \frac{1}{2}$, we have $t \geq 1$. Let $n \geq t+1$ and $D = \{x \in \{0,1\}^n : \ x_1 = 0, \ x_2 = \cdots = x_{t+1} = 1\}$, so that $\mu_{1/2}(D) = \eta \leq \frac{1}{4}$. Define

$$f_0 \equiv c, \qquad f_k \equiv z_k \ \ (k \notin S), \qquad f_k = z_k \oplus \mathbf{1}_D \ \ (k \in S).$$

The right-hand side of the test is $g(z \oplus \mathbf{1}_{\mathbf{S}})$ with $\mathbf{S} = \{k \in S : \mathbf{Z}^k \in D\}$. As (z, S) is a flip pair, it differs from c iff $\mathbf{S} = S$, so $\mathrm{err}_g(\mathbf{f}) = \eta^s \leq (2\varepsilon)^s$.

Let $\mathbf{F}$ be a multi-polymorphism of g; we show that some distance between $\mathbf{f}$ and $\mathbf{F}$ is at least $\min\{\eta, 2^{-m}\} > \varepsilon$. If g is affine, then $s = 1$, $p = \frac{1}{2}$ and $g = x_1 \oplus \cdots \oplus x_m \oplus b$. The proof of Lemma 8.14 for $m \geq 2$, applied to $\mathbf{F}$ (so $\delta = 0$), gives $|\hat{\Phi}_\ell(S_\ell)| = 1$ with a common set $S_\ell = I$; hence there are $I \subseteq [n]$ and $a_0, \ldots, a_m$ with $F_\ell = \bigoplus_{i\in I} x_i \oplus a_\ell$ for all ℓ. If $I \neq \emptyset$ then $\Pr_{\mu_{1/2}}[F_0 \neq c] = \frac{1}{2}$; if $I = \emptyset$, then for $k \in S$ the function F_k is constant and $\Pr[f_k \neq F_k] \geq \min\{\eta, 1-\eta\} = \eta$. If g is not affine, then by the classification of exact multi-polymorphisms [6, Theorem 2.5] (essentially due to Dokow and Holzman [8]), $\mathbf{F}$ is of one of the following types.

- *Certificate:* $F_0 \equiv c'$, and there are $J' \subseteq [m]$ and $b' \in \{0,1\}^{J'}$ with $F_j \equiv b'_j$ for $j \in J'$ and $g(x) = c'$ whenever $x_{J'} = b'$. If $c' \neq c$, then $\Pr_{\mu_p}[f_0 \neq F_0] = 1$. If $b'_j \neq z_j$ for some $j \in J'$, then $\Pr[f_j \neq F_j] \geq 1 - \eta$. Otherwise $g \equiv c$ on $\{x : x_{J'} = z_{J'}\}$; this set does not contain $z \oplus \mathbf{1}_S$, where $g \neq c$, so there is $j \in J' \cap S$, and $\Pr[f_j \neq F_j] = \Pr[z_j \oplus \mathbf{1}_D \neq z_j] = \eta$.

- *Dictator:* $F_0 = x_i \oplus a_0$ for some $i \in [n]$ and $a_0 \in \{0,1\}$. Then $\Pr_{\mu_p}[f_0 \neq F_0] \in \{p, 1-p\}$, and $\min\{p, 1-p\} \geq 2^{-m}$.

- *OR and AND type:* for some $I \subseteq [n]$, every F_j with $j \in [m]$ is OR_I or AND_I. For $j \in S$ the function f_j is $\mathbf{1}_D$ or $1 \oplus \mathbf{1}_D$, and the claim in the proof of Theorem 8.18(a) gives $\Pr[f_j \neq F_j] \geq \eta$.

So $\mathbf{f}$ is not ε-close to any multi-polymorphism, and $\delta_g^{\mathrm{multi}}(\varepsilon) \leq \eta^s \leq (2\varepsilon)^s$. □

Proof of Theorem 8.28(b),(c). Let $\mathbf{f}$ have $\mathrm{err}_g(\mathbf{f}) = \delta$, and put $E = E(g)$ and $p = p_g$. The dictator tuple $(x_1, \ldots, x_1)$ is a multi-polymorphism of every g at distance at most 1 from $\mathbf{f}$. So if the claim holds for $\delta < \delta_g$, for some $\delta_g > 0$, then it holds for all δ once $C_g \geq \delta_g^{-1/E}$.

Affine g. Here $E = 1$ by Lemma 8.27(c), and Lemma 8.14 gives distance at most δ when $\delta < \min\{1/50, 1/(m+2)\}$.

g of AND type. Let $g(x) = c \oplus \prod_k (x_k \oplus a_k)$, so $E = m$ and $p \in \{2^{-m}, 1-2^{-m}\}$. Put $\mathbf{W}^k = \mathbf{Z}^k \oplus a_k\mathbf{1}$, $h_k(w) = f_k(w \oplus a_k\mathbf{1}) \oplus a_k$ for $k \in [m]$, and $h_0(v) = f_0(v \oplus c\mathbf{1}) \oplus c$. Then $\mathbf{W}^1, \dots, \mathbf{W}^m$ are independent and uniform, $g(f_1(\mathbf{Z}^1), \dots, f_m(\mathbf{Z}^m)) = c \oplus \prod_k h_k(\mathbf{W}^k)$, and $g^n(\mathbf{Z}) = c\mathbf{1} \oplus (\mathbf{W}^1 \wedge \cdots \wedge \mathbf{W}^m)$, so $f_0(g^n(\mathbf{Z})) = c \oplus h_0(\mathbf{W}^1 \wedge \cdots \wedge \mathbf{W}^m)$. Thus the test for $\mathbf{f}$ at $\mathbf{Z}$ is the test of Lemma 8.31 for $\mathbf{h}$ at $\mathbf{W}$. Let $\mathbf{H}$ be the tuple given by Lemma 8.31, and let $F_k(x) = H_k(x \oplus a_k\mathbf{1}) \oplus a_k$ and $F_0(y) = H_0(y \oplus c\mathbf{1}) \oplus c$. The same computation shows that $\mathbf{F}$ is a multi-polymorphism of g. Moreover $\Pr[f_k \neq F_k] = \Pr[h_k \neq H_k]$, and $\Pr_{\mu_p}[f_0 \neq F_0] = \Pr_{\mu_{2^{-m}}}[h_0 \neq H_0]$ because $y \mapsto y \oplus c\mathbf{1}$ maps μ_p to $\mu_{2^{-m}}$. This proves (c), for every δ.

The remaining functions. From now on g is neither affine nor of AND type. (For $m = 2$ there are no such functions.)

Constants. For $\emptyset \neq S \subsetneq [m]$ let $K_S(y \mid x) = \Pr_{\mathbf{u}}[g(x, \mathbf{u}) = y]$ for $x \in \{0,1\}^S$, with $\mathbf{u}$ uniform on $\{0,1\}^{[m]\setminus S}$, as in Lemma 8.23, and let π_S be the uniform distribution on $\{0,1\}^S$. Then $\theta(\pi_S, K_S) < 1$: pick $k \notin S$; as g depends on coordinate k, some x has $K_S(0 \mid x) > 0$ and $K_S(1 \mid x) > 0$, and since every x is adjacent to 0 or to 1, the bipartite graph of Lemma 8.22(b) is connected. Let $\theta_* < 1$ be the maximum of $\theta(\pi_S, K_S)$ over these S; define M, θ_1, C_1 as in Lemma 8.30 with $\theta = \theta_*$ and $p = p_g$; and put $\kappa = 2\theta_1/(1+\theta_1) < 1$ and $\tau_1 = (1-\theta_1)/2$. For $0 < \gamma \leq \frac{1}{4}$ let

$$\alpha(\gamma) = (\gamma/2)^m, \qquad \pi_0(\gamma) = \left(2^{-m-3}\alpha(\gamma)^{2C_1}\right)^{1/(1-\kappa)},$$

$$L(\gamma) = 2^{m+2}\ln(2\gamma^{1-m}), \qquad K(\gamma) = (m+2)\,2^{2m+5}\gamma^{-3(m-1)}.$$

Let $\gamma_1 = \frac{1}{4}$ and $\gamma_{i+1} = \min\{\gamma_i/4,\ \pi_0(\gamma_i)\}$ for $i = 1, \dots, m+1$. Let C° and c_g be the constants of Theorem 8.1 for g, and let $\delta_g > 0$ be such that $C^\circ \delta_g^{c_g} \leq 3\gamma_{m+1}/8$ and $K(\gamma_{m+1})\delta_g < 2^{-mL(\gamma_{m+1})}$. Assume $\delta < \delta_g$.

Near-constant and far columns. For $k \in [m]$ let b_k be a majority value of f_k and $\eta_k = \Pr[f_k \neq b_k] = \min_b \Pr[f_k \neq b]$. The $m+1$ intervals $(\gamma_{i+1}, \gamma_i]$, $i \in [m+1]$, are disjoint, so one of them contains none of $\eta_1, \dots, \eta_m$. Fix such an i and put $\gamma = \gamma_i$ and $\gamma' = \gamma_{i+1}$, so that $\gamma \geq \gamma_{m+1}$, $\gamma' \leq \gamma/4$ and $\gamma' \leq \pi_0(\gamma)$. Let $J = \{k : \eta_k \leq \gamma'\}$ and $N = \{k : \eta_k > \gamma\}$. Then $[m] = J \sqcup N$, and f_k is γ-far from constants for $k \in N$.

Case A: $J = \emptyset$. Fix j^* with $a_{j^*} > 0$ (Lemma 8.4), so $a_{j^*} \geq 2^{1-m}$, and fix $v \in B_{j^*}$. Then $P_{j^*} \geq \prod_{k \neq j^*} \Pr[f_k = v_k] \geq \gamma^{m-1}$, and Lemma 8.9 gives $T \subseteq [n]$ with $|T| \leq L(\gamma)$ and $\beta_* := \Pr[f_{j^*} \neq \mathrm{rd}_T f_{j^*}] \leq 2^{m+2}\gamma^{1-m}\delta$. Lemma 8.10 gives $\beta_0 := \Pr_{\mu_p}[f_0 \neq \mathrm{rd}_T^p f_0] \leq 2^m(\delta + \beta_*) \leq 2^{2m+3}\gamma^{1-m}\delta$, and Lemma 8.11 gives $\beta_j := \Pr[f_j \neq \mathrm{rd}_T f_j] \leq 2(\delta + \beta_0)\gamma^{-2(m-1)} \leq 2^{2m+5}\gamma^{-3(m-1)}\delta$ for every $j \in [m]$. Hence $\delta + \beta_0 + \sum_j \beta_j \leq K(\gamma)\delta \leq K(\gamma_{m+1})\delta < 2^{-mL(\gamma_{m+1})} \leq 2^{-m|T|}$, and by Lemma 8.13 the tuple $(\mathrm{rd}_T^p f_0, \mathrm{rd}_T f_1, \dots, \mathrm{rd}_T f_m)$ is a multi-polymorphism, at distance at most $K(\gamma_{m+1})\delta \leq K(\gamma_{m+1})\delta^{1/E}$ from $\mathbf{f}$.

Case B: $J \neq \emptyset$. Let $C_J = \{x \in \{0,1\}^m : \ x_k = b_k \text{ for } k \in J\}$.

Case B2: g is not constant on C_J. Then $N \neq \emptyset$, since otherwise C_J is a single point. Theorem 8.1 gives a multi-polymorphism $\mathbf{F}$ at distance at most $\rho := C^\circ\delta^{c_g} \leq 3\gamma_{m+1}/8 \leq 3\gamma/8$ from $\mathbf{f}$. With $\eta(h) = \min_b \Pr[h \neq b]$ we have $|\eta_k - \eta(F_k)| \leq \rho$ for every k. As in the proof of (a), $\mathbf{F}$ is of one of three types, and each leads to a contradiction.

- Certificate (c', J', b'): for $j \in J'$, $\eta_j \leq \Pr[f_j \neq b'_j] \leq \rho < \gamma$, so $j \in J$; and $b'_j = b_j$, since otherwise $\Pr[f_j \neq b_j] \geq 1 - \rho > \frac{1}{2}$. Hence $C_J \subseteq \{x : x_{J'} = b'\}$, where $g \equiv c'$, contradicting the assumption of Case B2.

- Dictator: $\eta(F_k) = \frac{1}{2}$ for all $k \in [m]$, so $\eta_k \geq \frac{1}{2} - \rho > \frac{1}{4} \geq \gamma'$ for all k, contradicting $J \neq \emptyset$.

- OR or AND type with index set I: if $I = \emptyset$, all F_k are constant and $\eta_k \le \rho < \gamma$ for all k, contradicting $N \neq \emptyset$. If $I \neq \emptyset$, then $\eta(F_k) = 2^{-|I|}$ for all $k \in [m]$. Taking $j \in J$ and $k \in N$ gives $\gamma - \rho < \eta_k - \rho \le 2^{-|I|} \le \eta_j + \rho \le \gamma' + \rho$, so $\gamma - \gamma' < 2\rho \le \frac{3}{4}\gamma$, contradicting $\gamma' \le \gamma/4$. (In fact this type does not occur here: by [6, Theorem 2.5] it requires g to be an OR or an AND of m literals, i.e. of AND type.)

So Case B2 does not occur.

Case B1: $g \equiv c$ on C_J for some $c \in \{0,1\}$. For $S \subseteq J$ let $C_J^S = \{x : x_k = b_k \oplus \mathbf{1}\{k \in S\} \text{ for } k \in J\}$ (so $C_J^\emptyset = C_J$), and call S a *flip set* if g is not identically c on C_J^S. Let $\mathcal{M}$ be the family of inclusion-minimal flip sets. As g is not constant, $\mathcal{M} \neq \emptyset$, and $\emptyset \notin \mathcal{M}$. We use three facts.

(F1) *If $S \in \mathcal{M}$, then $|S| \le E$ and $S \neq [m]$.* Pick $x \in C_J^S$ with $g(x) \neq c$ and put $z = x \oplus \mathbf{1}_S \in C_J$, so $g(z) = c$. For $S' \subsetneq S$ we have $z \oplus \mathbf{1}_{S'} \in C_J^{S'}$, and S' is not a flip set, so $g(z \oplus \mathbf{1}_{S'}) = c$. Hence (z, S) is a flip pair, and $|S| \le E$. If $S = [m]$, then $E = m$ and g would be of AND type.

(F2) *If $J' \subseteq J$ meets every member of $\mathcal{M}$, then $\mathbf{F} = (F_0, \dots, F_m)$ with $F_0 \equiv c$, $F_k \equiv b_k$ for $k \in J'$ and $F_k = f_k$ for $k \notin J'$ is a multi-polymorphism.* Let $x \in \{0,1\}^m$ with $x_k = b_k$ for $k \in J'$, and let $S = \{k \in J : x_k \neq b_k\}$, so $x \in C_J^S$ and $S \cap J' = \emptyset$. If $g(x) \neq c$, then S is a flip set and contains a member of $\mathcal{M}$, which meets J'; this is impossible. So $g(x) = c$, and both sides of the test for $\mathbf{F}$ always equal c.

(F3) For $k \in J$ let $A_k = \{x : f_k(x) \neq b_k\}$, so $\mu_{1/2}(A_k) = \eta_k$, and for $S \subseteq J$ let $\pi'_S = \prod_{k \in S} \eta_k$. The right-hand side of the test for $\mathbf{f}$ is $g(x)$ with $x = (f_k(\mathbf{Z}^k))_k \in C_J^{\mathbf{S}}$, where $\mathbf{S} = \{k \in J : \mathbf{Z}^k \in A_k\}$. If it differs from c, then $\mathbf{S}$ is a flip set and contains some $S \in \mathcal{M}$. By the independence of the columns, $\Pr[\text{RHS} \neq c] \le \sum_{S \in \mathcal{M}} \pi'_S \le 2^m \pi^*$, where $\pi^* = \max_{S \in \mathcal{M}} \pi'_S$.

Let $B = \{y : f_0(y) \neq c\}$ and $\nu_B = \mu_p(B)$. The test fails whenever $g^n(\mathbf{Z}) \in B$ and RHS $= c$, and $g^n(\mathbf{Z}) \sim \mu_p$; so by (F3),

$$\delta \;\ge\; \nu_B - 2^m \pi^*. \tag{8}$$

Claim: $\delta \ge c_* \pi^*$, where $c_* = \tau_1 \alpha(\gamma_{m+1})$. Let $S \in \mathcal{M}$ with $\pi'_S = \pi^*$; we may assume $\pi^* > 0$. By (F1), $\emptyset \neq S \subsetneq [m]$. Fix $x^* \in C_J^S$ with $g(x^*) \neq c$. Let $\mathcal{E}_1$ be the event that $\mathbf{Z}^k \in A_k$ for all $k \in S$, and $\mathcal{E}_2$ the event that $\mathbf{Z}^k \notin A_k$ for all $k \in J \setminus S$ and $f_k(\mathbf{Z}^k) = x^*_k$ for all $k \in N$. They are independent, $\Pr[\mathcal{E}_1] = \pi^*$, and $\Pr[\mathcal{E}_2] \ge (3/4)^{|J \setminus S|} \gamma^{|N|} \ge \alpha(\gamma) \ge \alpha(\gamma_{m+1})$, because $\eta_k \le \gamma' \le \frac{1}{4}$ for $k \in J$ and $\Pr[f_k = b] > \gamma$ for $k \in N$ and $b \in \{0,1\}$. On $\mathcal{E}_1 \cap \mathcal{E}_2$ we have $(f_k(\mathbf{Z}^k))_k = x^*$, so the right-hand side is $g(x^*) \neq c$, and the test fails if $g^n(\mathbf{Z}) \notin B$. Let P be the conditional law of $g^n(\mathbf{Z})$ given $\mathcal{E}_1 \cap \mathcal{E}_2$. Then

$$\delta \;\ge\; \pi^* \alpha(\gamma)\,(1 - P(B)). \tag{9}$$

To bound $P(B)$, write the rows as $\mathbf{z}_i = (\mathbf{x}_i, \mathbf{u}_i)$ with $\mathbf{x}_i \in \{0,1\}^S$ and $\mathbf{u}_i \in \{0,1\}^{[m] \setminus S}$. The event $\mathcal{E}_1$ depends only on $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ and $\mathcal{E}_2$ only on $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_n)$, so given $\mathcal{E}_1 \cap \mathcal{E}_2$ the blocks $\mathbf{x}$ and $\mathbf{u}$ are independent, with laws $\rho_X = P_{\mathbf{x}|\mathcal{E}_1}$ and $\rho_U = P_{\mathbf{u}|\mathcal{E}_2}$, and $D(\rho_X \| \pi_S^{\otimes n}) = \ln(1/\pi^*)$ and $D(\rho_U \| \pi_{[m]\setminus S}^{\otimes n}) = \ln(1/\Pr[\mathcal{E}_2]) \le \ln(1/\alpha(\gamma))$. Lemma 8.30, with $h = g$, $\mathcal{X} = \{0,1\}^S$, $\mathcal{U} = \{0,1\}^{[m]\setminus S}$, $K = K_S$ and θ_* in place of $\theta(\pi_S, K_S)$, gives

$$D(P \| \mu_p) \;\le\; \theta_1 \ln \frac{1}{\pi^*} + C_1 \ln \frac{1}{\alpha(\gamma)}.$$

By data processing for the map $y \mapsto \mathbf{1}_B(y)$, and since $d(u\|v) \ge u \ln \frac{1}{v} - \ln 2$, we get $P(B) \ln \frac{1}{\nu_B} \le \theta_1 \ln \frac{1}{\pi^*} + C_1 \ln \frac{1}{\alpha(\gamma)} + \ln 2$ (if $\nu_B = 0$ then $B = \emptyset$ and $P(B) = 0$). If $P(B) \le 1 - \tau_1$, then (9) gives

$\delta \geq \tau_1 \alpha(\gamma)\pi^* \geq c_*\pi^*$. If $P(B) > 1 - \tau_1 = (1+\theta_1)/2$, then

$$\ln \frac{1}{\nu_B} < \kappa \ln \frac{1}{\pi^*} + \frac{2}{1+\theta_1}\Big(C_1 \ln \frac{1}{\alpha(\gamma)} + \ln 2\Big) \leq \kappa \ln \frac{1}{\pi^*} + 2C_1 \ln \frac{1}{\alpha(\gamma)} + 2\ln 2,$$

i.e. $\nu_B > (\pi^*)^\kappa \alpha(\gamma)^{2C_1}/4$. As $\pi^* \leq \min_{k\in S} \eta_k \leq \gamma' \leq \pi_0(\gamma)$, we have $(\pi^*)^{1-\kappa} \leq 2^{-m-3}\alpha(\gamma)^{2C_1}$, so $\nu_B > 2^{m+1}\pi^* \geq (2^m+1)\pi^*$, and (8) gives $\delta \geq \pi^* \geq c_*\pi^*$. This proves the claim.

The certificate. Let $t = \max_{S\in\mathcal{M}} \min_{k\in S} \eta_k$, attained at $S_1 \in \mathcal{M}$, and $J' = \{k \in J : \eta_k \leq t\}$. Every $S \in \mathcal{M}$ contains some k with $\eta_k \leq t$, so J' meets every member of $\mathcal{M}$, and by (F2) the tuple $\mathbf{F}$ defined there is a multi-polymorphism. Its distances to $\mathbf{f}$ are $\eta_k \leq t$ for $k \in J'$, 0 for $k \notin J'$, and ν_B for f_0. Now $t^{|S_1|} \leq \pi'_{S_1} \leq \pi^* \leq \delta/c_*$ and $|S_1| \leq E$ by (F1), so $t \leq (\delta/c_*)^{1/E} \leq c_*^{-1}\delta^{1/E}$ (if $\delta/c_* \geq 1$ this holds because $t \leq 1$). By (8) and the claim, $\nu_B \leq \delta + 2^m\pi^* \leq (1 + 2^m/c_*)\delta$. So $\mathbf{F}$ is $((2^m+1)/c_*)\,\delta^{1/E}$-close to $\mathbf{f}$.

Conclusion. In all cases $\mathbf{f}$ is $C_g\delta^{1/E}$-close to a multi-polymorphism, with

$$C_g = \max\big\{\delta_g^{-1/E},\ K(\gamma_{m+1}),\ (2^m+1)/c_*\big\}$$

for the remaining functions, $C_g = \max\{50, m+2\}$ for affine g, and $C_g = 2m+11$ for g of AND type. Hence every $(\varepsilon/C_g)^E$-approximate multi-polymorphism is ε-close to a multi-polymorphism, i.e. $\delta_g^{\mathrm{multi}}(\varepsilon) \geq (\varepsilon/C_g)^E$, and together with (a), $\delta_g^{\mathrm{multi}}(\varepsilon) = \Theta_g(\varepsilon^E)$. As in Theorem 8.18, the only non-explicit quantity in C_g (for the remaining functions) is θ_*, which comes from the compactness argument of Lemma 8.22(b). If $\lambda_S = \sum_y \min_x K_S(y \mid x) > 0$ for every $\emptyset \neq S \subsetneq [m]$, then $\theta(\pi_S, K_S) \leq 1 - \lambda_S$ by Lemma 8.22(c), and by the last sentence of the proof of Lemma 8.30 one may take $\theta_* = \max_S(1-\lambda_S)$, which makes C_g explicit. □

Remark 8.32. For a single function the near-constant case is controlled by the two constant inputs, and balanced g have linear loss. For tuples, near-constant columns can be combined with columns far from constants, the relevant inputs are all points of all restrictions, and $E(g)$ can exceed both $e(g)$ and the polymorphism exponent. By exact enumeration (with the script `multi_exponent_census.py` in the directory `numerics`), among the functions that depend on all their m coordinates the values of E are distributed as follows: for $m = 3$ (218 functions), $E = 1$: 2, $E = 2$: 200, $E = 3$: 16; for $m = 4$ (64,594 functions), $E = 1$: 2, $E = 2$: 43,104, $E = 3$: 21,456, $E = 4$: 32. The exponent $E(g)$ exceeds the optimal exponent for polymorphisms (1 if $p_g = 1/2$, $e(g)$ otherwise) for 184 of the 218 functions with $m = 3$ and for 62,492 of the 64,594 functions with $m = 4$. The same script checks Lemma 8.27(b),(c), the characterization of the functions of AND type, and the description of E by restrictions, for all g with $m \leq 4$.

8.8 Correlated inputs: every full-support distribution

Setting. Let μ be a distribution on $\{0,1\}^m$ with full support, and put $\alpha = \min_z \mu(z) > 0$. In this subsection the rows $\mathbf{z}_1, \ldots, \mathbf{z}_n$ of $\mathbf{Z}$ are independent samples of μ. Then the column $\mathbf{Z}^j$ has law μ_{p_j}, where $p_j = \Pr_\mu[\mathbf{z}_j = 1]$, and $g^n(\mathbf{Z})$ has law μ_{p_g}, where now $p_g = \Pr_\mu[g(\mathbf{z}) = 1]$; but different columns are correlated. Let

$$\mathrm{err}_g^\mu(\mathbf{f}) = \Pr\big[f_0(g^n(\mathbf{Z})) \neq g\big(f_1(\mathbf{Z}^1), \ldots, f_m(\mathbf{Z}^m)\big)\big].$$

Distances to f_0 are measured under μ_{p_g} and distances to f_j under μ_{p_j}; "β-close to a constant" and "majority value" refer to the same measures. Since every matrix has positive probability, $\mathrm{err}_g^\mu(\mathbf{F}) = 0$ if and only if $\mathbf{F}$ is a multi-polymorphism of g; this notion does not depend on μ. Uniform μ gives the setting of Theorems 8.1 and 8.28.

This is the setting of Alekseev and Filmus [2] for functional predicates. The map $z \mapsto (z, g(z))$ identifies the full-support distributions on $\{0,1\}^m$ with the full-support distributions on $P_g = \{(z, g(z)) : z \in \{0,1\}^m\} \subseteq \{0,1\}^{m+1}$, and $(f_1, \dots, f_m, f_0)$ is a (μ, δ)-approximate generalized polymorphism of P_g in their sense iff $\mathrm{err}_g^{\mu}(\mathbf{f}) \le \delta$. In their main theorem [2, Theorem 1.6] (every predicate, every full-support distribution) closeness is measured with respect to the marginals of μ, as here, and the dependence of δ on ε is of tower type. They conjecture that "the optimal dependence is polynomial or even linear" [2, Section 9], and they prove linear dependence for the predicates $P_{m,b} = \{a \in \{0,1\}^m : a_1 \oplus \cdots \oplus a_m = b\}$, $m \ge 3$ [2, Theorem 1.7]. Theorem 1.4(c) (the FLMM setting $\mu_p \times \mu_\rho$) and Theorems 8.1–8.28 only concern product distributions. We prove the polynomial part of the conjecture for every functional predicate.

For $\kappa \in (0,1]$ let $\Gamma_1(\kappa) = 1$ and $\Gamma_k(\kappa) = \frac{2}{\kappa}\big(1 + \Gamma_{k-1}(\kappa)\big)$ for $k \ge 2$; by induction (using $1 \le (2/\kappa)^{k-2}$), $\Gamma_k(\kappa) \le k(2/\kappa)^{k-1}$. For $m \ge 2$ and $0 < \alpha \le 1/4$ put

$$c_{m,\alpha} = \frac{\alpha^3}{8\,\Gamma_m(\alpha)\big((\ell_\alpha + 3)\Gamma_{m-1}(\alpha) + 1\big)}, \qquad \ell_\alpha = \frac{16}{\alpha} \ln \frac{1}{\alpha}.$$

Since $\Gamma_k(\alpha)$ and ℓ_α decrease in α on $(0, 1/4]$ and $\Gamma_k(\alpha)$ increases in k, $c_{m,\alpha}$ is increasing in α and decreasing in m.

Theorem 8.33 (Polynomial dependence under correlated inputs)**.** *Let $g\colon \{0,1\}^m \to \{0,1\}$, and let μ be a distribution on $\{0,1\}^m$ with full support and $\alpha = \min_z \mu(z)$. There is $C = C(g, \mu)$ such that for every $n \ge 1$ and every tuple $\mathbf{f}$ with $\mathrm{err}_g^{\mu}(\mathbf{f}) = \delta$ there is a multi-polymorphism $\mathbf{F}$ of g with*

$$\Pr_{\mu_{p_g}}[f_0 \ne F_0] \le C\delta^c \qquad \text{and} \qquad \Pr_{\mu_{p_j}}[f_j \ne F_j] \le C\delta^c \quad (j \in [m]),$$

where $c = 1$ if g is affine over $\mathbb{F}_2$ (including constant g), and $c = c_{m,\alpha}$ otherwise. Moreover, if g is not affine, $\mathbf{F}$ can be chosen with $F_\ell = F_{\ell'}$ whenever $f_\ell = f_{\ell'}$ and $p_\ell = p_{\ell'}$ $(\ell, \ell' \in \{0, \dots, m\}$, $p_0 := p_g)$.

For affine g with $m \ge 2$ the theorem is [2, Theorem 1.7]; the new part is the polynomial bound for every non-affine g. Thus, for functional predicates, δ can be taken polynomial in ε in [2, Theorem 1.6]. The exponent $c_{m,\alpha}$ is far from optimal; for uniform μ, Theorems 8.1, 8.18 and 8.28 give much better (and in part optimal) exponents.

Corollary 8.34 (AND testing with correlated inputs)**.** *Let μ be a distribution on $\{0,1\}^2$ with full support and equal marginals $p = \mu(1,0) + \mu(1,1) = \mu(0,1) + \mu(1,1)$, let $\alpha = \min \mu$, and let $(\mathbf{x}, \mathbf{y}) \in (\{0,1\}^n)^2$ have independent coordinate pairs $(\mathbf{x}_i, \mathbf{y}_i) \sim \mu$. If $\varepsilon = \Pr[f(\mathbf{x} \wedge \mathbf{y}) \ne f(\mathbf{x}) f(\mathbf{y})]$, then there is $h \in \mathcal{A}_n$ with $\Pr_{\mu_p}[f \ne h] \le C\varepsilon^c$ and $\Pr_{\mu_{\mu(1,1)}}[f \ne h] \le C\varepsilon^c$, where $c = c_{2,\alpha} = \alpha^4/(32(\ell_\alpha + 4))$ and $C = C(\mu)$.*

Proof. Apply Theorem 8.33 to $g = \mathrm{AND}_2$ and $f_0 = f_1 = f_2 = f$. Since $p_1 = p_2$, we get $F_1 = F_2 =: h$, and $F_0(x \wedge y) = h(x)h(y)$ for all x, y. Taking $y = x$ gives $F_0 = h$, so $h(x \wedge y) = h(x)h(y)$, i.e. $h \in \mathcal{A}_n$. Finally $p_g = \mu(1,1)$, and $c_{2,\alpha} = \alpha^3/(8 \cdot \frac{4}{\alpha}(\ell_\alpha + 4))$. □

Reverse hypercontractivity. The proofs of Lemmas 8.8–8.12 use the independence of the columns. Under a correlated μ we replace it by the reverse hypercontractive inequality of Mossel, Oleszkiewicz and Sen [22], which bounds from below, polynomially, the probability that two correlated events occur together. Reverse hypercontractivity was used before in exactly this role, for coordinates that are correlated within each row, by Mossel [23] and in [22, Section 1, Application 2] to prove quantitative Arrow theorems under general full-support product

distributions. There the dependence of δ on ε is explicit but not polynomial: [22, (10.1)] gives $\delta = \exp\big(-C\alpha^{-7}2^{\alpha^{-2}}\varepsilon^{-2-1/(2\alpha^2)}\log^2(1/\varepsilon)\big)$, where α is the smallest atom of the law of the vector of pairwise comparisons within a triple of alternatives (for three alternatives, the least probability of a ranking). Moreover, the Arrow predicate is not functional. For $\kappa \in (0,1)$ let

$$r_\kappa = \frac{2(1-\sqrt{1-\kappa})}{2-\sqrt{1-\kappa}} \in \Big[\frac{8\kappa}{9}, \kappa\Big),$$

since, with $t = \sqrt{1-\kappa} \in (0,1)$ and $\kappa = (1-t)(1+t)$, we have $r_\kappa/\kappa = 2/\big((2-t)(1+t)\big) = 2/\big(2+t(1-t)\big)$ and $0 < t(1-t) \le \frac{1}{4}$. For a nonnegative random variable V and $r \in (0,1)$ write $\|V\|_r = (\mathbb{E}V^r)^{1/r}$.

Lemma 8.35 (Reverse hypercontractivity)**.** *Let $(\mathbf{x}_i, \mathbf{y}_i)$, $i \in [n]$, be independent pairs of random variables with finitely many values, and let $\kappa \in (0,1)$ be such that $\Pr[\mathbf{x}_i = a, \mathbf{y}_i = b] \ge \kappa \Pr[\mathbf{x}_i = a]\Pr[\mathbf{y}_i = b]$ for all i, a, b. Put $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ and $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$. Then for all nonnegative functions Φ, Ψ,*

$$\mathbb{E}[\Phi(\mathbf{x})\Psi(\mathbf{y})] \;\ge\; \|\Phi(\mathbf{x})\|_{r_\kappa}\,\|\Psi(\mathbf{y})\|_{r_\kappa}.$$

In particular $\Pr[\mathbf{x} \in A,\ \mathbf{y} \in B] \ge (\Pr[\mathbf{x} \in A]\Pr[\mathbf{y} \in B])^{1/r_\kappa} \ge (\Pr[\mathbf{x} \in A]\Pr[\mathbf{y} \in B])^{2/\kappa}$ for all sets A, B.

For identically distributed pairs, this set version is [22, Lemma 8.3] (stated there with $\min(\Pr[\mathbf{x} \in A], \Pr[\mathbf{y} \in B])$ in place of the geometric mean); the general form follows from their Corollary 8.2 by the induction below.

Proof. Let $n = 1$. If $\mathbf{x}_1$ and $\mathbf{y}_1$ are independent, then $\mathbb{E}[\Phi(\mathbf{x}_1)\Psi(\mathbf{y}_1)] = \mathbb{E}[\Phi(\mathbf{x}_1)]\,\mathbb{E}[\Psi(\mathbf{y}_1)] \ge \|\Phi(\mathbf{x}_1)\|_{r_\kappa}\|\Psi(\mathbf{y}_1)\|_{r_\kappa}$, since $\|V\|_r \le \mathbb{E}V$ for $r \in (0,1)$ by Jensen's inequality. Otherwise we use [22, Corollary 8.2] with $n = 1$ and $p = q = r_\kappa$ (numbering of arXiv:1108.1210). Embed the ranges of $\mathbf{x}_1$ and $\mathbf{y}_1$ in one finite set Ω, let μ and ν be the laws of $\mathbf{x}_1$ and $\mathbf{y}_1$, and let $K(a,\cdot)$ be the conditional law of $\mathbf{y}_1$ given $\mathbf{x}_1 = a$ if $\mu(a) > 0$, and ν otherwise. Then $\nu = \mu K$, and the quantity $\alpha_K = \min_{a,b:\nu(b)>0} K(a,b)/\nu(b)$ of [22, (8.1)] satisfies $\kappa \le \alpha_K < 1$: the first inequality is our hypothesis, and $\alpha_K = 1$ would give $K(a,b) \ge \nu(b)$ for all a, b, hence $K(a,\cdot) = \nu$ for every a (both sum to 1), i.e. independence. So $\alpha_K^\star = -\log(1-\alpha_K)$ is finite, the decomposition $K = T_{\alpha_K^\star} S$ in the proof of [22, Proposition 8.1] is defined (MOS13 do not treat $\alpha_K = 1$, where $\alpha_K^\star = \infty$), and the condition $\alpha_K^\star \ge \log\frac{(2-p)(2-q)}{4(1-p)(1-q)}$ of [22, Corollary 8.2] holds, because for $p = q = r_\kappa$ its right-hand side equals $-\log(1-\kappa) \le \alpha_K^\star$, as $\frac{2-r_\kappa}{2(1-r_\kappa)} = (1-\kappa)^{-1/2}$. The conclusion $\mathbb{E}_\mu[\Phi \cdot K\Psi] \ge \|\Phi\|_{L^{r_\kappa}(\mu)}\|\Psi\|_{L^{r_\kappa}(\nu)}$ is our claim for $n = 1$.

For $n \ge 2$ we induct on n; write $r = r_\kappa$. The vectors $\mathbf{x}' = (\mathbf{x}_1, \dots, \mathbf{x}_{n-1})$ and $\mathbf{y}' = (\mathbf{y}_1, \dots, \mathbf{y}_{n-1})$ are jointly independent of $(\mathbf{x}_n, \mathbf{y}_n)$. For fixed values a, b of $\mathbf{x}_n, \mathbf{y}_n$ the induction hypothesis gives $\mathbb{E}[\Phi(\mathbf{x}', a)\Psi(\mathbf{y}', b)] \ge \phi(a)\psi(b)$, where $\phi(a) = \|\Phi(\mathbf{x}', a)\|_r$ and $\psi(b) = \|\Psi(\mathbf{y}', b)\|_r$. Hence, by the case $n = 1$,

$$\mathbb{E}[\Phi(\mathbf{x})\Psi(\mathbf{y})] \;\ge\; \mathbb{E}[\phi(\mathbf{x}_n)\psi(\mathbf{y}_n)] \;\ge\; \|\phi(\mathbf{x}_n)\|_r\,\|\psi(\mathbf{y}_n)\|_r \;=\; \|\Phi(\mathbf{x})\|_r\,\|\Psi(\mathbf{y})\|_r,$$

where the last step uses $\mathbb{E}[\phi(\mathbf{x}_n)^r] = \mathbb{E}[\Phi(\mathbf{x})^r]$, which holds because $\mathbf{x}'$ and $\mathbf{x}_n$ are independent, and similarly for ψ. For indicators $\|\mathbf{1}_A(\mathbf{x})\|_r = \Pr[\mathbf{x} \in A]^{1/r}$, and $1/r_\kappa \le 2/\kappa$. □

Lemma 8.36 (Several columns)**.** *Let ν be a distribution on $\{0,1\}^k$, $k \ge 1$, with $\nu(z) \ge \kappa$ for all z, where $\kappa \in (0,1)$, and let $\mathbf{Y}$ be an $n \times k$ matrix whose rows are independent samples of ν. If $A_1, \dots, A_k \subseteq \{0,1\}^n$ and $\Pr[\mathbf{Y}^\ell \in A_\ell] \ge \eta$ for every $\ell \in [k]$, then $\Pr[\mathbf{Y}^\ell \in A_\ell \text{ for all } \ell \in [k]] \ge \eta^{\Gamma_k(\kappa)}$.*

Proof. Induction on k; the case $k=1$ is trivial. For $k \geq 2$ let $\mathbf{x}_i$ be the first entry of row i of $\mathbf{Y}$ and $\mathbf{y}_i$ the vector of its other $k-1$ entries. Then $\Pr[\mathbf{x}_i = a, \mathbf{y}_i = b] = \nu(a,b) \geq \kappa \geq \kappa \Pr[\mathbf{x}_i = a] \Pr[\mathbf{y}_i = b]$. With B the event that $\mathbf{Y}^\ell \in A_\ell$ for $2 \leq \ell \leq k$, Lemma 8.35 shows that the probability in question is at least $(\eta \Pr[B])^{2/\kappa}$. The rows of $(\mathbf{Y}^2, \dots, \mathbf{Y}^k)$ are independent samples of the marginal of ν on the last $k-1$ coordinates, whose values are at least $2\kappa \geq \kappa$, so $\Pr[B] \geq \eta^{\Gamma_{k-1}(\kappa)}$ by induction. Hence the probability is at least $\eta^{\frac{2}{\kappa}(1+\Gamma_{k-1}(\kappa))} = \eta^{\Gamma_k(\kappa)}$. □

Lemma 8.37 (One resampled coordinate, correlated)**.** *Let g depend on coordinate j, let $\mathbf{z} \sim \mu$, and let $\mathbf{z}'$ be obtained from $\mathbf{z}$ by replacing $\mathbf{z}_j$ with a fresh sample of the conditional law of $\mathbf{z}_j$ given $\mathbf{z}_{-j}$. Then $\mathbf{z}' \sim \mu$, and $(g(\mathbf{z}), g(\mathbf{z}'))$ is an exchangeable pair of Bernoulli(p_g) bits with $\Pr[g(\mathbf{z}) \neq g(\mathbf{z}')] = 2p_g(1-p_g)(1-\rho_j)$ for some $\rho_j \in [0,1)$ with $\frac{1}{1-\rho_j} \leq \frac{1}{4\alpha^2}$.*

Proof. For a context u let $\mu_{-j}(u) = \mu(u^{j\to 0}) + \mu(u^{j\to 1})$ and $q(u) = \mu(u^{j\to 1})/\mu_{-j}(u)$. Given $\mathbf{z}_{-j} = u$, the bits $\mathbf{z}_j, \mathbf{z}'_j$ are independent Bernoulli$(q(u))$ bits. So $(\mathbf{z}, \mathbf{z}')$ is exchangeable, $\mathbf{z}' \sim \mu$, and given $\mathbf{z}_{-j}$ the bits $g(\mathbf{z}), g(\mathbf{z}')$ are independent and identically distributed. As in Lemma 8.7, $\Pr[g(\mathbf{z}) = g(\mathbf{z}') = 1] = \mathbb{E}[r(\mathbf{z}_{-j})^2] \geq p_g^2$ with $r(u) = \Pr[g(\mathbf{z}) = 1 \mid \mathbf{z}_{-j} = u]$, so $\rho_j \geq 0$ (note $0 < p_g < 1$, as g is not constant and μ has full support). Since $B_j \neq \emptyset$ and $\mu_{-j} \leq 1$,

$$\Pr[g(\mathbf{z}) \neq g(\mathbf{z}')] = \sum_{u \in B_j} \mu_{-j}(u)\, 2q(u)(1-q(u)) = \sum_{u \in B_j} \frac{2\mu(u^{j\to 0})\,\mu(u^{j\to 1})}{\mu_{-j}(u)} \;\geq\; 2\alpha^2 > 0.$$

Hence $\rho_j < 1$ and $\frac{1}{1-\rho_j} = \frac{2p_g(1-p_g)}{\Pr[g(\mathbf{z}) \neq g(\mathbf{z}')]} \leq \frac{1/2}{2\alpha^2}$. □

As before, $\mathbf{c}_j = (f_k(\mathbf{Z}^k))_{k \neq j}$, and A_j, B_j are the sets of insensitive and sensitive contexts for j (they depend only on g). Write $\delta = \mathrm{err}_g^\mu(\mathbf{f})$ and $\mathrm{RHS}(\mathbf{Z}) = g(f_1(\mathbf{Z}^1), \dots, f_m(\mathbf{Z}^m))$.

Lemma 8.38 (Junta lemma, correlated)**.** *Let $j \in [m]$ with $A_j \neq \emptyset$, suppose $P_j := \Pr[\mathbf{c}_j \in B_j] > 0$, and let $\hat{f}_j$ be the Fourier coefficients of f_j with respect to μ_{p_j}. Then there is $T \subseteq [n]$ with*

$$\textit{(a)}\quad |T| \leq \frac{16}{\alpha} \ln \frac{2}{P_j}, \qquad\qquad \textit{(b)}\quad \Pr_{\mu_{p_j}}[f_j \neq \mathrm{rd}_T^{p_j} f_j] \leq 2 \sum_{S \not\subseteq T} \hat{f}_j(S)^2 \leq \frac{8\,(2\delta)^{\alpha/4}}{\alpha P_j}.$$

Proof. A latent decomposition. Fix $u^* \in A_j$, and let ν_1 be the law of the random vector with context u^* and an independent Bernoulli(p_j) bit in position j. Put $\sigma = \alpha/2$. Since $\sigma\nu_1 \leq \alpha/2$ pointwise, $\nu_0 = (\mu - \sigma\nu_1)/(1-\sigma)$ is a probability distribution with $\nu_0(z) \geq \mu(z) - \alpha/2 \geq \alpha/2$ for every z, and its j-th marginal is Bernoulli(p_j), like those of μ and ν_1. Let $\zeta_1, \dots, \zeta_n \in \{0,1\}$ be independent with $\Pr[\zeta_i = 1] = \sigma$, and given $\zeta = (\zeta_1, \dots, \zeta_n)$ let the rows $\mathbf{z}_i$ be independent with laws ν_{ζ_i}. The rows are then independent samples of $\sigma\nu_1 + (1-\sigma)\nu_0 = \mu$, so $\mathbf{Z}$ has the right law. Let $\mathbf{R} = \{i : \zeta_i = 1\}$. Given ζ, the entries of $\mathbf{Z}^j$ are independent Bernoulli(p_j) bits.

Containment. Let $\mathbf{W}$ be obtained from $\mathbf{Z}$ by replacing the entries $\mathbf{Z}_{ij}$, $i \in \mathbf{R}$, by fresh independent Bernoulli(p_j) bits. Given ζ the rows of $\mathbf{W}$ are independent with laws ν_{ζ_i}, so $\mathbf{W}$ has the law of $\mathbf{Z}$. The columns $k \neq j$ are unchanged, so $\mathbf{c}_j(\mathbf{W}) = \mathbf{c}_j(\mathbf{Z})$, and $g^n(\mathbf{W}) = g^n(\mathbf{Z})$ because the rows in $\mathbf{R}$ have the insensitive context u^*. As in the proof of Lemma 8.8, the event $E = \{\mathbf{c}_j \in B_j,\ f_j(\mathbf{Z}^j) \neq f_j(\mathbf{W}^j)\}$ forces the test to fail on $\mathbf{Z}$ or on $\mathbf{W}$, so $\Pr[E] \leq 2\delta$.

Reverse hypercontractivity. Condition on ζ. The pairs $(\mathbf{x}_i, \mathbf{y}_i)$ with $\mathbf{x}_i = (\mathbf{Z}_{ij}, \mathbf{W}_{ij})$ and $\mathbf{y}_i = (\mathbf{z}_i)_{-j}$, $i \in [n]$, are conditionally independent. For $i \in \mathbf{R}$, $\mathbf{y}_i = u^*$ is constant, so $\mathbf{x}_i$ and $\mathbf{y}_i$ are independent. For $i \notin \mathbf{R}$, $\mathbf{x}_i = (\mathbf{Z}_{ij}, \mathbf{Z}_{ij})$ and $\Pr[\mathbf{x}_i = (a,a),\ \mathbf{y}_i = u \mid \zeta] = \nu_0(u^{j\to a}) \geq \alpha/2$.

So Lemma 8.35 applies conditionally on ζ, with $\kappa = \alpha/2$ and $r = r_{\alpha/2}$, to the indicators of $\{f_j(\mathbf{Z}^j) \neq f_j(\mathbf{W}^j)\}$ and $\{\mathbf{c}_j \in B_j\}$:

$$\Pr[E \mid \zeta] \;\geq\; \big(\Pr[f_j(\mathbf{Z}^j) \neq f_j(\mathbf{W}^j) \mid \zeta] \cdot \omega(\zeta)\big)^{1/r}, \qquad \omega(\zeta) = \Pr[\mathbf{c}_j \in B_j \mid \zeta].$$

Given ζ, $(\mathbf{Z}^j, \mathbf{W}^j)$ is the pair of Lemma 2.1 with $p = p_j$ and $J = [n] \setminus \mathbf{R}$, so the first factor equals $2W(\mathbf{R})$, where $W(R) = \sum_{S \cap R \neq \emptyset} \hat{f}_j(S)^2$. Taking expectations and using Jensen's inequality for the convex function $t \mapsto t^{1/r}$ on $[0,\infty)$,

$$2\delta \;\geq\; \mathbb{E}\big[(2W(\mathbf{R})\,\omega(\zeta))^{1/r}\big] \;\geq\; \big(\mathbb{E}[2W(\mathbf{R})\,\omega(\zeta)]\big)^{1/r}, \quad \text{so}$$
$$\sum_S \hat{f}_j(S)^2\, \mathbb{E}\big[\omega(\zeta)\mathbf{1}\{S \cap \mathbf{R} \neq \emptyset\}\big] \;\leq\; \frac{(2\delta)^r}{2}.$$

Note that $\mathbb{E}[\omega(\zeta)] = P_j$.

The junta. Let $\pi_i = \mathbb{E}[\omega(\zeta)\mathbf{1}\{i \in \mathbf{R}\}]$ and $T = \{i \in [n] : \pi_i \leq \sigma P_j/4\}$.

(b) If $S \not\subseteq T$, pick $i \in S \setminus T$; then $\mathbb{E}[\omega(\zeta)\mathbf{1}\{S \cap \mathbf{R} \neq \emptyset\}] \geq \pi_i > \sigma P_j/4$. Hence $\sum_{S \not\subseteq T} \hat{f}_j(S)^2 \leq 2(2\delta)^r/(\sigma P_j) = 4(2\delta)^r/(\alpha P_j)$, and Lemma 8.5 gives the first inequality of (b). If $2\delta \leq 1$, then $(2\delta)^r \leq (2\delta)^{\alpha/4}$ because $r \geq \alpha/4$. If $2\delta > 1$, the right-hand side of (b) exceeds $8/\alpha > 2$, and (b) holds trivially.

(a) We may assume $T \neq \emptyset$. The ζ_i are independent, so $\mathbf{N} = |\mathbf{R} \cap T|$ is binomial with mean $\sigma|T|$, and the Chernoff bound gives $\Pr[\mathbf{N} \leq \sigma|T|/2] \leq e^{-\sigma|T|/8}$. On the other hand $\mathbb{E}[\omega(\zeta)\mathbf{N}] = \sum_{i \in T} \pi_i \leq \sigma|T|P_j/4$, so by Markov's inequality $\mathbb{E}[\omega(\zeta)\mathbf{1}\{\mathbf{N} \geq \sigma|T|/2\}] \leq P_j/2$, i.e. $\mathbb{E}[\omega(\zeta)\mathbf{1}\{\mathbf{N} < \sigma|T|/2\}] \geq P_j/2$. Since $\omega \leq 1$, we get $P_j/2 \leq \Pr[\mathbf{N} < \sigma|T|/2] \leq e^{-\sigma|T|/8}$, i.e. $|T| \leq \frac{8}{\sigma} \ln \frac{2}{P_j}$. □

Lemma 8.39 (From one column to f_0, correlated)**.** *Let g depend on coordinate j, let $T \subseteq [n]$, and suppose $\Pr_{\mu_{p_j}}[f_j \neq J] \leq \beta$ for some T-junta J. Then*

$$\Pr_{\mu_{p_g}}[f_0 \neq \mathrm{rd}_T^{p_g} f_0] \;\leq\; \frac{2\delta + 2\beta}{1 - \rho_j} \;\leq\; \frac{\delta + \beta}{2\alpha^2}.$$

Proof. Let $\mathbf{W}$ be obtained from $\mathbf{Z}$ by replacing, independently for each $i \notin T$, the entry $\mathbf{Z}_{ij}$ by a fresh sample of the conditional law of $\mathbf{z}_j$ given that the context is $(\mathbf{z}_i)_{-j}$. By Lemma 8.37 the rows of $\mathbf{W}$ are independent samples of μ. The columns $k \neq j$ are unchanged, and $\mathbf{W}^j \sim \mu_{p_j}$ agrees with $\mathbf{Z}^j$ on T, so $J(\mathbf{W}^j) = J(\mathbf{Z}^j)$ and $\Pr[f_j(\mathbf{Z}^j) \neq f_j(\mathbf{W}^j)] \leq 2\beta$. Hence $\mathbf{y} = g^n(\mathbf{Z})$ and $\mathbf{y}' = g^n(\mathbf{W})$ satisfy $\Pr[f_0(\mathbf{y}) \neq f_0(\mathbf{y}')] \leq 2\delta + 2\beta$. Moreover $\mathbf{y}_T = \mathbf{y}'_T$, and the pairs $(\mathbf{y}_i, \mathbf{y}'_i)$, $i \notin T$, are i.i.d., independent of $\mathbf{y}_T$, with the law of Lemma 8.37. The end of the proof of Lemma 8.10 (Lemma 8.6 with $q = p_g$ and $\rho = \rho_j$, for each value of $\mathbf{y}_T$) now applies verbatim. □

Lemma 8.40 (From f_0 to every column, correlated)**.** *Let $T \subseteq [n]$, suppose $\Pr_{\mu_{p_g}}[f_0 \neq J_0] \leq \beta$ for some T-junta J_0, let g depend on coordinate j, and suppose that f_k is γ-far from constants for every $k \in [m] \setminus \{j\}$. Then*

$$\Pr_{\mu_{p_j}}[f_j \neq \mathrm{rd}_T^{p_j} f_j] \;\leq\; (2\delta + 2\beta)^{\alpha^2/2}\, \gamma^{-2\Gamma_{m-1}(\alpha)}.$$

Proof. Let $\mathbf{W}$ be obtained from $\mathbf{Z}$ by replacing every row $\mathbf{z}_i$, $i \notin T$, by a fresh independent sample of μ, and let $\mathbf{w}_i$ be the rows of $\mathbf{W}$. As in the proof of Lemma 8.11, $\mathbf{W}$ has the law of $\mathbf{Z}$ and $\Pr[\mathrm{RHS}(\mathbf{Z}) \neq \mathrm{RHS}(\mathbf{W})] \leq 2\delta + 2\beta$. Fix $v \in B_j$, let A be the event that $f_k(\mathbf{Z}^k) = v_k$ for all $k \neq j$, and let E be the event that $f_k(\mathbf{Z}^k) = f_k(\mathbf{W}^k) = v_k$ for all $k \neq j$.

$\Pr[E] \ge \gamma^{2\Gamma_{m-1}(\alpha)}$. Condition on the entries $\mathbf{Z}_{ik}$ with $i \in T$ and $k \neq j$. Given them, the families $(\mathbf{Z}_{ik})_{i\notin T, k\neq j}$ and $(\mathbf{W}_{ik})_{i\notin T, k\neq j}$ are independent and identically distributed, while $\mathbf{W}_{ik} = \mathbf{Z}_{ik}$ for $i \in T$. So the conditional probability of E is the square of that of A, and $\Pr[E] \ge \Pr[A]^2$ by the Cauchy–Schwarz inequality. The columns $\mathbf{Z}^k$, $k \neq j$, have independent rows whose law is the marginal of μ on the coordinates $k \neq j$, with values at least $2\alpha \ge \alpha$, and $\Pr[f_k(\mathbf{Z}^k) = v_k] > \gamma$ for $k \neq j$. So Lemma 8.36 gives $\Pr[A] \ge \gamma^{\Gamma_{m-1}(\alpha)}$.

Reverse hypercontractivity. Let $\mathbf{x}_i = (\mathbf{Z}_{ij}, \mathbf{W}_{ij})$ and $\mathbf{y}_i = ((\mathbf{z}_i)_{-j}, (\mathbf{w}_i)_{-j})$; these pairs are independent over i. For $i \in T$ we have $\mathbf{w}_i = \mathbf{z}_i$ and $\Pr[\mathbf{x}_i = (a,a),\ \mathbf{y}_i = (u,u)] = \mu(u^{j\to a}) \ge \alpha$. For $i \notin T$, $\Pr[\mathbf{x}_i = (a,a'),\ \mathbf{y}_i = (u,u')] = \mu(u^{j\to a})\mu(u'^{j\to a'}) \ge \alpha^2$. So Lemma 8.35 applies with $\kappa = \alpha^2$, and with $D = \{f_j(\mathbf{Z}^j) \neq f_j(\mathbf{W}^j)\}$ it gives $\Pr[D \cap E] \ge (\Pr[D]\Pr[E])^{2/\alpha^2}$. On E the two right-hand sides are $g(v^{j\to f_j(\mathbf{Z}^j)})$ and $g(v^{j\to f_j(\mathbf{W}^j)})$, which differ iff D occurs, because v is sensitive. Hence $\Pr[D \cap E] \le 2\delta + 2\beta$, and

$$\Pr[D] \ \le\ \frac{(2\delta+2\beta)^{\alpha^2/2}}{\Pr[E]} \ \le\ (2\delta+2\beta)^{\alpha^2/2}\,\gamma^{-2\Gamma_{m-1}(\alpha)}.$$

Since $\mathbf{W}^j$ agrees with $\mathbf{Z}^j$ on T and is a fresh sample of μ_{p_j} off T, Lemma 8.5 finishes the proof. □

Lemma 8.41 (Certificates, correlated)**.** *Let g depend on all its m coordinates, let $0 < \varepsilon \le 1/2$, $\varepsilon' = \alpha^2\varepsilon^{\Gamma_m(\alpha)}/4$ and $\delta \le \varepsilon'$, and suppose that some f_{k_0}, $k_0 \in [m]$, is ε'-close to a constant. Define J, the majority values a_j $(j \in J)$ and a_0, and the tuple $\mathbf{F}$ as in Lemma 8.12 (with the measures of this subsection). Then $\mathbf{F}$ is a multi-polymorphism of g, and all its distances to $\mathbf{f}$ are at most ε.*

Proof. A constant is a $\emptyset$-junta, so Lemma 8.39 with $T = \emptyset$ and $j = k_0$ gives $\Pr_{\mu_{p_g}}[f_0 \neq a_0] \le (\delta + \varepsilon')/(2\alpha^2) \le \varepsilon'/\alpha^2 = \varepsilon^{\Gamma_m(\alpha)}/4 \le \varepsilon$. Fix $x \in \{0,1\}^m$ with $x_J = a_J$. For every $k \in [m]$ we have $\Pr[f_k(\mathbf{Z}^k) = x_k] \ge \varepsilon$: for $k \in J$ because $x_k = a_k$ is the majority value, and for $k \notin J$ because f_k is ε-far from constants. By Lemma 8.36 (with $\nu = \mu$ and $\kappa = \alpha$), $\mathrm{RHS}(\mathbf{Z}) = g(x)$ with probability at least $\varepsilon^{\Gamma_m(\alpha)}$. On the other hand, $\mathrm{RHS}(\mathbf{Z}) \neq a_0$ with probability at most $\delta + \Pr_{\mu_{p_g}}[f_0 \neq a_0] \le \varepsilon' + \varepsilon^{\Gamma_m(\alpha)}/4 \le \varepsilon^{\Gamma_m(\alpha)}/2$. Hence $g(x) = a_0$, and the proof ends as that of Lemma 8.12. □

Lemma 8.42 (Exactness on a junta, correlated)**.** *Let $T \subseteq [n]$ and let $F_0, \dots, F_m$ be T-juntas with $\delta + \Pr_{\mu_{p_g}}[f_0 \neq F_0] + \sum_{j=1}^m \Pr_{\mu_{p_j}}[f_j \neq F_j] < \alpha^{|T|}$. Then $\mathbf{F}$ is a multi-polymorphism of g.*

Proof. By the union bound $\mathrm{err}_g^{\mu}(\mathbf{F}) < \alpha^{|T|}$. Whether $F_0(g^n(\mathbf{Z})) = \mathrm{RHS}_{\mathbf{F}}(\mathbf{Z})$ holds depends only on the rows $\mathbf{z}_i$, $i \in T$, and each value of these rows has probability at least $\alpha^{|T|}$. So the identity holds for every value of these rows, i.e. for every $\mathbf{Z}$. □

Proof of Theorem 8.33. Reductions. If $\delta = 0$ take $\mathbf{F} = \mathbf{f}$. The all-dictator tuple $F_0 = \cdots = F_m = (x \mapsto x_1)$ is a multi-polymorphism of every g, since both sides of the test equal $g(\mathbf{z}_1)$; it satisfies the "moreover" part, and its distances to $\mathbf{f}$ are at most $1 \le \delta_0^{-c}\delta^c$ whenever $\delta \ge \delta_0$. So it suffices to treat $0 < \delta < \delta_0$ for some $\delta_0 = \delta_0(g,\mu) > 0$, provided $C \ge \delta_0^{-c}$. Let $\mathcal{D} \subseteq [m]$ be the set of coordinates on which g depends, G the induced function, and $\mu_{\mathcal{D}}$ the marginal of μ on $\{0,1\}^{\mathcal{D}}$; it has full support, $\min \mu_{\mathcal{D}} \ge \alpha$, $p_G = p_g$, and $\mathrm{err}_g^{\mu}(\mathbf{f}) = \mathrm{err}_G^{\mu_{\mathcal{D}}}(f_0, (f_j)_{j\in\mathcal{D}})$. A multi-polymorphism of G, extended by arbitrary F_j for $j \notin \mathcal{D}$, is a multi-polymorphism of g. For $j \notin \mathcal{D}$ we take $F_j = F_\ell$ if $f_j = f_\ell$ and $p_j = p_\ell$ for some $\ell \in \mathcal{D} \cup \{0\}$, and $F_j = f_j$ otherwise; this keeps the distances and the "moreover" part. If G is not affine then $|\mathcal{D}| \ge 2$, and $c_{|\mathcal{D}|,\min\mu_{\mathcal{D}}} \ge c_{m,\alpha}$ by the monotonicity of $c_{m,\alpha}$. So we may assume that g depends on all its coordinates. If $m = 0$, then $g \equiv b$ and $\mathrm{err}_g^{\mu}(\mathbf{f}) = \mathbf{1}\{f_0(b\mathbf{1}) \neq b\}$, so $\delta < 1$ means that $\mathbf{f}$ is exact.

Affine g. If $m = 1$ and $g(x) = x \oplus b$, take $F_1 = f_1$ and $F_0(y) = f_1(y \oplus b\mathbf{1}) \oplus b$; this is exact, and $\Pr_{\mu_{p_g}}[f_0 \neq F_0] = \delta$ because $\mathbf{Z}^1 \oplus b\mathbf{1} \sim \mu_{p_g}$. If $m \geq 2$ and $g(x) = x_1 \oplus \cdots \oplus x_m \oplus b$, then $P_g = P_{m+1,b}$, the law of $(\mathbf{z}, g(\mathbf{z}))$ for $\mathbf{z} \sim \mu$ is a full-support distribution on $P_{m+1,b}$ whose marginals are those of μ and Bernoulli(p_g), and $(f_1, \ldots, f_m, f_0)$ is a δ-approximate generalized polymorphism of $P_{m+1,b}$. Since $m + 1 \geq 3$, [2, Theorem 1.7] gives a generalized polymorphism of $P_{m+1,b}$, i.e. a multi-polymorphism of g, within distance $O_\mu(\delta)$.

Non-affine g. Then $m \geq 2$ and $\alpha \leq 2^{-m} \leq 1/4$; fix j^* with $A_{j^*} \neq \emptyset$ (Lemma 8.4). Write $\Gamma = \Gamma_{m-1}(\alpha)$, $\Gamma' = \Gamma_m(\alpha)$, $c = c_{m,\alpha}$, and

$$K = \frac{8\big((\ell_\alpha + 3)\Gamma + 1\big)}{\alpha^3} = \frac{1}{c\,\Gamma'}, \qquad \varepsilon_0 = \frac{2^{-\ell_\alpha}\alpha^3}{20(m+2)}.$$

Given δ, define ε' and ε by $\delta = \varepsilon'^K$ and $\varepsilon' = \alpha^2\varepsilon^{\Gamma'}/4$, so that $\varepsilon = (4/\alpha^2)^{1/\Gamma'}\delta^c$. Choose δ_0 such that $\delta < \delta_0$ implies $\varepsilon \leq 1/2$ and $\varepsilon' < \varepsilon_0$. Then $\delta \leq \varepsilon' \leq \varepsilon$.

Case I: some f_k, $k \in [m]$, is ε'-close to a constant. Lemma 8.41 gives a multi-polymorphism within distance ε; here the majority values a_j ($j \in J$) and a_0 are $\mathrm{rd}_\emptyset^{p_j} f_j$ and $\mathrm{rd}_\emptyset^{p_g} f_0$, with ties broken towards 1.

Case II: every f_k, $k \in [m]$, is ε'-far from constants. Fix $v \in B_{j^*}$. Lemma 8.36 (applied to the columns $k \neq j^*$, whose rows are independent samples of the marginal of μ off j^*, with values $\geq \alpha$) gives $P_{j^*} \geq \Pr[\mathbf{c}_{j^*} = v] \geq \varepsilon'^\Gamma$. Let T be the set of Lemma 8.38 for $j = j^*$. Then, using $(2\delta)^{\alpha/4} \leq 2\delta^{\alpha/4}$,

$$|T| \leq \frac{16}{\alpha}\Big(\ln 2 + \Gamma \ln \frac{1}{\varepsilon'}\Big), \qquad \beta_{j^*} := \Pr_{\mu_{p_{j^*}}}\left[f_{j^*} \neq \mathrm{rd}_T^{p_{j^*}} f_{j^*}\right] \leq \frac{8(2\delta)^{\alpha/4}}{\alpha\,\varepsilon'^\Gamma} \leq \frac{16}{\alpha}\,\varepsilon'^{\alpha K/4-\Gamma}.$$

Lemma 8.39 (with $J = \mathrm{rd}_T^{p_{j^*}} f_{j^*}$), together with $\delta \leq \varepsilon'^{\alpha K/4-\Gamma}$, gives $\beta_0 := \Pr_{\mu_{p_g}}[f_0 \neq \mathrm{rd}_T^{p_g} f_0] \leq (\delta + \beta_{j^*})/(2\alpha^2) \leq \frac{9}{\alpha^3}\varepsilon'^{\alpha K/4-\Gamma}$. So $2\delta + 2\beta_0 \leq \frac{20}{\alpha^3}\varepsilon'^{\alpha K/4-\Gamma}$, and Lemma 8.40 (with $J_0 = \mathrm{rd}_T^{p_g} f_0$ and $\gamma = \varepsilon'$) gives, for every $j \in [m]$,

$$\beta_j := \Pr_{\mu_{p_j}}[f_j \neq \mathrm{rd}_T^{p_j} f_j] \;\leq\; \Big(\frac{20}{\alpha^3}\Big)^{\alpha^2/2} \varepsilon'^{\frac{\alpha^2}{2}(\frac{\alpha K}{4}-\Gamma)-2\Gamma} \;\leq\; \frac{20}{\alpha^3}\,\varepsilon'^{\frac{\alpha^3 K}{8}-3\Gamma} \;=\; \frac{20}{\alpha^3}\,\varepsilon'^{\ell_\alpha\Gamma+1} =: D.$$

Also $\delta \leq D$ and $\beta_0 \leq D$, since $K \geq \alpha K/4 - \Gamma \geq \ell_\alpha\Gamma + 1$. As $|T| \ln \frac{1}{\alpha} \leq \ell_\alpha(\ln 2 + \Gamma \ln \frac{1}{\varepsilon'})$, we have $\alpha^{|T|} \geq 2^{-\ell_\alpha}\varepsilon'^{\ell_\alpha\Gamma}$, and since $\varepsilon' < \varepsilon_0$,

$$\delta + \beta_0 + \sum_{j=1}^m \beta_j \;\leq\; (m+2)D \;=\; \frac{20(m+2)}{\alpha^3}\,\varepsilon' \cdot \varepsilon'^{\ell_\alpha\Gamma} \;<\; 2^{-\ell_\alpha}\varepsilon'^{\ell_\alpha\Gamma} \;\leq\; \alpha^{|T|}.$$

By Lemma 8.42, $\mathbf{F} = (\mathrm{rd}_T^{p_g} f_0, \mathrm{rd}_T^{p_1} f_1, \ldots, \mathrm{rd}_T^{p_m} f_m)$ is a multi-polymorphism of g, and its distances to $\mathbf{f}$ are at most $D < \varepsilon' \leq \varepsilon$.

In both cases all distances are at most $\varepsilon \leq (4/\alpha^2)\,\delta^c$, which proves the theorem with $C = \max(4/\alpha^2, \delta_0^{-c})$.

The "moreover" part. In Case II, $\mathrm{rd}_T^{p_\ell} f_\ell$ depends only on f_ℓ, p_ℓ and T. In Case I, J and the a_j are determined by the pairs (f_j, p_j); and if $f_0 = f_j$ and $p_g = p_j$ for some $j \in [m]$, then f_j is ε-close to a_0 under μ_{p_j}, so $j \in J$, and $a_j = \mathrm{rd}_\emptyset^{p_j} f_j = \mathrm{rd}_\emptyset^{p_g} f_0 = a_0$. □

Remark 8.43 (Dependence on μ). For product μ, Lemma 8.36 is trivial, and the reverse hypercontractive steps can be replaced by independence, as in Lemmas 8.9 and 8.11. In our proof the exponent depends on α through the reverse hypercontractive steps and through the threshold $\alpha^{|T|}$

in Lemma 8.42. Some dependence on μ is necessary, and the exponent must tend to 0 as $\min\mu \to 0$, already for AND testing:

Claim. Let μ be a full-support distribution on $\{0,1\}^2$ with equal marginals p, put $a = \mu(1,1)$ and $\sigma^2 = p(1-p)$, and let $\rho = (a-p^2)/\sigma^2 \in (-1,1)$ be the correlation of the two coordinates. If Corollary 8.34, or Theorem 8.33 for $g = \mathrm{AND}_2$, holds for μ with an exponent c, then $c \le \min\{\frac{1}{2}, \frac{1+\rho}{2}\}$.

For example, $\mu(0,0) = \mu(1,1) = \alpha \le \frac{1}{4}$ and $\mu(0,1) = \mu(1,0) = \frac{1}{2} - \alpha$ give $\min\mu = \alpha$, $p = \frac{1}{2}$, $\rho = 4\alpha - 1$, and hence $c \le 2\alpha$; this case was pointed out by a referee. For uniform μ the claim gives $c \le \frac{1}{2}$, which is attained (Theorem 1.1, and Theorem 8.28 with $E(\mathrm{AND}_2) = 2$). Our exponent $c_{2,\alpha} = \alpha^4/(32(\ell_\alpha + 4))$ is of order $\alpha^5/\log(1/\alpha)$. For $\rho > 0$ we do not know whether $\frac{1}{2}$ is attained. In that case $\mu = \gamma\,\delta_{(1,1)} + (1-\gamma)\,\mu_{p'} \otimes \mu_{p'}$ with $p' = p(1-\rho)$ and $\gamma = \rho p/(1-p+\rho p)$, so $(\mathbf{x},\mathbf{y})$ can be generated as $(\mathbf{u} \vee \mathbf{1}_{\mathbf{E}}, \mathbf{v} \vee \mathbf{1}_{\mathbf{E}})$, where $\mathbf{E}$ contains each coordinate independently with probability γ and $\mathbf{u},\mathbf{v} \sim \mu_{p'}$ are independent. Since $\mathbf{x} \wedge \mathbf{y} = (\mathbf{u} \wedge \mathbf{v}) \vee \mathbf{1}_{\mathbf{E}}$, the μ-test is the independent p'-biased test applied to the upper restriction $u \mapsto f(u \vee \mathbf{1}_{\mathbf{E}})$, so by Theorem 5.1 each upper restriction is close to an AND or to 0. We do not know how to glue these approximations with exponent $\frac{1}{2}$.

Proof of the claim: $c \le (1+\rho)/2$. Fix $t > 0$. For $n \ge 1$ let $k = \lceil pn + t\sqrt{n} \rceil$, $f(x) = \mathbf{1}\{|x| \ge k\}$ and $f_0 = f_1 = f_2 = f$, and let $\varepsilon_n = \Pr[f(\mathbf{x} \wedge \mathbf{y}) \ne f(\mathbf{x})f(\mathbf{y})]$, which is also $\mathrm{err}^{\mu}_{\mathrm{AND}_2}(f,f,f)$. If the test fails, then $|\mathbf{x} \wedge \mathbf{y}| \ge k$ or $|\mathbf{x}|, |\mathbf{y}| \ge k$. Now $|\mathbf{x} \wedge \mathbf{y}|$ is binomial with parameters n and $a < p$ (as $\mu(1,0) > 0$), so $\Pr[|\mathbf{x} \wedge \mathbf{y}| \ge k] \le e^{-2n(p-a)^2}$ by Hoeffding's inequality. If $|\mathbf{x}|, |\mathbf{y}| \ge k$, then $\sum_i X_i \ge 2k - 2pn \ge 2t\sqrt{n}$ for the i.i.d. centred variables $X_i = \mathbf{x}_i + \mathbf{y}_i - 2p$, which satisfy $|X_i| \le 2$ and $\mathbb{E}X_i^2 = 2\sigma^2(1+\rho)$. By Bernstein's inequality this has probability at most

$$\exp\Big(-\frac{(2t\sqrt{n})^2/2}{2\sigma^2(1+\rho)n + \frac{2}{3}\cdot 2t\sqrt{n}}\Big) = \exp\Big(-\frac{t^2}{\sigma^2(1+\rho) + 2t/(3\sqrt{n})}\Big).$$

Hence $\limsup_{n\to\infty} \varepsilon_n \le e^{-t^2/(\sigma^2(1+\rho))}$.

Next, every multi-polymorphism of AND_2 either has $F_0 \equiv 0$ and ($F_1 \equiv 0$ or $F_2 \equiv 0$), or has $F_0 = F_1 = F_2 \in \mathcal{A}_n \setminus \{0\}$ (this classification is also stated in [2, Section 9]). Indeed, $F_0(x) = F_1(x)F_2(\mathbf{1})$ and $F_0(y) = F_1(\mathbf{1})F_2(y)$ for all x, y. If $F_1(\mathbf{1})F_2(\mathbf{1}) = 0$, then $F_0 \equiv 0$, so $F_1(x)F_2(y) = 0$ for all x, y. Otherwise $F_0 = F_1 = F_2$ is not identically zero and satisfies $F_0(x \wedge y) = F_0(x)F_0(y)$. Let $P_n = \Pr_{\mu_p}[f = 1]$ and $d_n = \min\{P_n,\ 1 - P_n,\ (1-p)\Pr[\mathrm{Bin}(n-1,p) \ge k]\}$. If $F_j \equiv 0$ for some $j \in \{1,2\}$, then $\Pr_{\mu_p}[f_j \ne F_j] = P_n$. If $F_1 = \mathrm{AND}_\emptyset \equiv 1$, then $\Pr_{\mu_p}[f_1 \ne F_1] = 1 - P_n$. If $F_1 = \mathrm{AND}_S$ with $i \in S$, then $\Pr_{\mu_p}[f_1 \ne F_1] \ge \Pr_{\mu_p}[f = 1,\ x_i = 0] = (1-p)\Pr[\mathrm{Bin}(n-1,p) \ge k]$. So some $j \in \{1,2\}$ has $\Pr_{\mu_p}[f_j \ne F_j] \ge d_n$, and likewise $\Pr_{\mu_p}[f \ne h] \ge d_n$ for every $h \in \mathcal{A}_n$. By the central limit theorem, P_n and $\Pr[\mathrm{Bin}(n-1,p) \ge k]$ tend to $\bar{\Phi}(t/\sigma) < \frac{1}{2}$, where $\bar{\Phi}$ is the standard normal tail, so $\lim_n d_n = (1-p)\bar{\Phi}(t/\sigma)$.

If the exponent c works with a constant C, then $d_n \le C\varepsilon_n^c$ for all n, and letting $n \to \infty$ gives $(1-p)\bar{\Phi}(t/\sigma) \le Ce^{-ct^2/(\sigma^2(1+\rho))}$ for every $t > 0$. Since $\bar{\Phi}(s) \ge \frac{s}{1+s^2}\cdot\frac{e^{-s^2/2}}{\sqrt{2\pi}}$ for $s > 0$, taking logarithms, dividing by t^2 and letting $t \to \infty$ gives $\frac{c}{\sigma^2(1+\rho)} \le \frac{1}{2\sigma^2}$, i.e. $c \le (1+\rho)/2$.

Proof of the claim: $c \le \frac{1}{2}$. Let $0 < q \le \frac{1}{8}$, $k = \lceil pn \rceil$ and $T = \{x : |x| \ge k\}$, and let $\mathbf{A} \subseteq T$ be a random set containing each $x \in T$ independently with probability q. Put $b = \mu(1,0) = p - a > 0$ and $\theta = \max(p, 1-p) < 1$. If $|x \wedge y| < k$, then $f = \mathbf{1}_{\mathbf{A}}$ vanishes at $x \wedge y$, so the test fails only if $|\mathbf{x} \wedge \mathbf{y}| \ge k$ or $\mathbf{x}, \mathbf{y} \in \mathbf{A}$. Since $\Pr[x, y \in \mathbf{A}] = q^2$ for $x \ne y$ and $\Pr[\mathbf{x} = \mathbf{y}] = (1-2b)^n$, Hoeffding's inequality (as in the first part) gives

$$\mathbb{E}_{\mathbf{A}}\,\varepsilon(\mathbf{1}_{\mathbf{A}}) \ \le\ e^{-2nb^2} + q^2 + q(1-2b)^n.$$

Let $\beta_n = (1-p)\Pr[\mathrm{Bin}(n-1,p) \ge k]$. Then $\mathbb{E}_{\mathbf{A}}\Pr_{\mu_p}[x \in \mathbf{A},\ x_i = 0] = q\Pr_{\mu_p}[x \in T,\ x_i = 0] = q\beta_n$ for every $i \in [n]$, and $\beta_n \to (1-p)/2$ by the central limit theorem. Each of the $n+1$ random

variables $\mu_p(\mathbf{A})$ and $\Pr_{\mu_p}[x \in \mathbf{A},\ x_i = 0]$ ($i \in [n]$) has the form $\sum_{x \in T} w_x \mathbf{1}\{x \in \mathbf{A}\}$ with fixed weights $0 \le w_x \le \mu_p(x) \le \theta^n$, so $\sum_x w_x^2 \le \theta^n$, and by Hoeffding's inequality it deviates from its mean by at least $q\beta_n/2$ with probability at most $2\exp(-q^2\beta_n^2\theta^{-n}/2)$. By Markov's inequality, $\varepsilon(\mathbf{1}_{\mathbf{A}}) > 3\mathbb{E}_{\mathbf{A}}\varepsilon(\mathbf{1}_{\mathbf{A}})$ with probability at most $\frac{1}{3}$. Hence, for all large n, some $A_n \subseteq T$ satisfies $\varepsilon(\mathbf{1}_{A_n}) \le 3(e^{-2nb^2} + q^2 + q(1-2b)^n)$, $\Pr_{\mu_p}[x \in A_n,\ x_i = 0] \ge q\beta_n/2$ for all i, and $\mu_p(A_n) \le q + q\beta_n/2 \le 2q$.

The first part, applied to $f = \mathbf{1}_{A_n}$ in place of the threshold function, shows that every multi-polymorphism $\mathbf{F}$ of AND_2 has $\Pr_{\mu_p}[f_j \ne F_j] \ge \min\{\mu_p(A_n), 1-\mu_p(A_n), \min_i \Pr_{\mu_p}[x \in A_n, x_i = 0]\}$ for some $j \in \{1,2\}$, and that $\Pr_{\mu_p}[\mathbf{1}_{A_n} \ne h]$ is at least the same minimum for every $h \in \mathcal{A}_n$. This minimum is at least $q\beta_n/2$, because $\mu_p(A_n) \ge \Pr_{\mu_p}[x \in A_n, x_1 = 0]$ and $1 - \mu_p(A_n) \ge 1 - 2q \ge \frac{3}{4}$. If the exponent c works with a constant C, then $q\beta_n/2 \le C\,\varepsilon(\mathbf{1}_{A_n})^c$, and letting $n \to \infty$ gives $q(1-p)/4 \le C(3q^2)^c$ for every $q \in (0, \frac{1}{8}]$. Letting $q \to 0$ gives $c \le \frac{1}{2}$. □

The script `alpha_dependence.py` (directory `numerics`) computes ε_n and d_n exactly for several μ, t and $n \le 3600$ and checks the bound on ε_n used above. The script `correlated_check.py` checks Lemmas 8.35, 8.36, 8.37, the main inequalities in the proof of Lemma 8.38 (the identity for $\Pr[f_j(\mathbf{Z}^j) \ne f_j(\mathbf{W}^j) \mid \zeta]$, the conditional reverse hypercontractive bound, $\Pr[E] \le 2\delta$, and (a), (b)), and the two inequalities in the proof of Lemma 8.40, by exact enumeration on small random instances ($n \le 4$ for $m = 2$, $n \le 3$ for $m = 3$).

Remark 8.44 (Equal functions, equal approximations). Alekseev and Filmus ask [2, Section 9, second question] whether their theorem can also guarantee $F_\ell = F_{\ell'}$ whenever $f_\ell = f_{\ell'}$ and the corresponding marginals of μ coincide. For every functional predicate P_g, $g\colon \{0,1\}^m \to \{0,1\}$, and every full-support μ, Theorem 8.33 gives this with polynomial dependence. For non-affine g it is the "moreover" part. For affine g it can be arranged as follows (we may assume g depends on all coordinates, as in the reductions above, and $\delta < \delta_0$, since otherwise the all-dictator tuple $F_0 = \cdots = F_m = (x \mapsto x_1)$, which is a multi-polymorphism of every g, is within distance $1 \le \delta_0^{-c}\delta^c$). If $m = 0$ there is nothing to prove. If $m = 1$ and $g(x) = x$, the tuple $F_0 = F_1 = f_1$ is exact and within distance δ. If $m = 1$ and $g(x) = x \oplus 1$, the constraint only matters when $p_1 = p_g = 1/2$ and $f_0 = f_1 =: f$; then choose a set H containing exactly one point of each pair $\{y, y \oplus \mathbf{1}\}$, and put $F(y) = f(y)$ for $y \in H$ and $F(y \oplus \mathbf{1}) = f(y) \oplus 1$ for $y \in H$. Then $F_0 = F_1 = F$ is exact, and $\Pr[F \ne f] \le \Pr[f(\mathbf{z} \oplus \mathbf{1}) \ne f(\mathbf{z}) \oplus 1] = \delta$ for uniform $\mathbf{z}$. If $m \ge 2$ and $g(x) = x_1 \oplus \cdots \oplus x_m \oplus b$, every multi-polymorphism has the form $F_k = \chi_S \oplus b_k$ ($0 \le k \le m$) for a common $S \subseteq [n]$, where $\chi_S(x) = \bigoplus_{i \in S} x_i$ (the classical solution of the Pexider equation over $\mathbb{F}_2^n$; we include the short proof). Indeed, setting all arguments except $x_k, x_{k'}$ to 0 and then also $x_{k'} = 0$ shows $F_k(x) = F_0(x \oplus b\mathbf{1}) \oplus c_k$ for every $k \in [m]$; and then $h(x) = F_0(x \oplus b\mathbf{1}) \oplus F_0(b\mathbf{1})$ satisfies $h(x \oplus x') = h(x) \oplus h(x')$, so $h = \chi_S$ for some S. Two such functions are equal or complementary. If $f_\ell = f_{\ell'}$ and $p_\ell = p_{\ell'}$, the tuple of [2, Theorem 1.7] satisfies $\Pr_{\mu_{p_\ell}}[F_\ell \ne F_{\ell'}] \le 2C\delta < 1$ once $\delta_0 < 1/(2C)$, so $F_\ell = F_{\ell'}$.

8.9 Larger alphabets: functional predicates

Setting. Let Σ be a finite set with $q = |\Sigma| \ge 2$, let $m \ge 2$, let $g\colon \Sigma^m \to \Sigma$, and let μ be a distribution on Σ^m with full support and $\alpha = \min_z \mu(z)$; note $\alpha \le q^{-m} \le 1/4$. As in Section 8.8, the rows $\mathbf{z}_1, \dots, \mathbf{z}_n$ of the matrix $\mathbf{Z} \in \Sigma^{n \times m}$ are independent samples of μ, $\mathbf{Z}^j \in \Sigma^n$ is the j-th column, and $\mathbf{Y} = g^n(\mathbf{Z})$, i.e. $\mathbf{Y}_i = g(\mathbf{z}_i)$. Let μ_j be the law of $\mathbf{z}_j$ for $\mathbf{z} \sim \mu$ ($j \in [m]$), let $\Sigma_0 = g(\Sigma^m)$, and let μ_0 be the law of $g(\mathbf{z})$, a full-support distribution on Σ_0. For $f_0, \dots, f_m\colon \Sigma^n \to \Sigma$ let

$$\mathrm{err}_g^{\mu}(\mathbf{f}) = \Pr\big[f_0(\mathbf{Y}) \ne \mathrm{RHS}(\mathbf{Z})\big], \qquad \mathrm{RHS}(\mathbf{Z}) = g\big(f_1(\mathbf{Z}^1), \dots, f_m(\mathbf{Z}^m)\big).$$

Distances to f_j are measured under $\mu_j^{\otimes n}$, and distances to f_0 under $\mu_0^{\otimes n}$ (only the values of f_0 on Σ_0^n matter). A *multi-polymorphism* is a tuple $\mathbf{F}$ with $F_0(g^n(Z)) = g(F_1(Z^1), \dots, F_m(Z^m))$ for

all $Z \in \Sigma^{n\times m}$; as μ has full support, this holds iff $\mathrm{err}_g^{\mu}(\mathbf{F}) = 0$. For $q = 2$ this is the setting of Section 8.8. Exactly as there, $(f_1, \dots, f_m, f_0)$ is a (μ, δ)-approximate generalized polymorphism of the predicate $P_g = \{(z, g(z))\} \subseteq \Sigma^{m+1}$ in the sense of [2] iff $\mathrm{err}_g^{\mu}(\mathbf{f}) \le \delta$.

Alekseev and Filmus conjecture that their main theorem extends to all finite alphabets, and prove it under the condition that every coordinate of the predicate is *flexible*, i.e. some $w \in P$ stays in P when that coordinate is changed arbitrarily [2, Theorem 1.10]; extending this to all predicates is their first open question [2, Section 9]. The last coordinate of P_g is never flexible, since changing $g(z)$ leaves P_g. So Theorem 1.10 of [2] covers no functional predicate (and it is stated without an explicit dependence of δ on ε). We prove a polynomial bound for every functional predicate satisfying the following condition.

For $j \in [m]$, a context $u \in \Sigma^{[m]\setminus\{j\}}$ and $a \in \Sigma$ let $u^{j\to a} \in \Sigma^m$ be u with a inserted in position j, and let z_{-j} be the context of $z \in \Sigma^m$. Let N_j, the *collision graph* of coordinate j, be the graph on Σ in which $a \neq b$ are adjacent iff $g(u^{j\to a}) = g(u^{j\to b})$ for some context u. Let $\mathcal{P}_j$ be the partition of Σ into the connected components of N_j, write $[a]_j \in \mathcal{P}_j$ for the component of a, and $a \approx_j b$ if $[a]_j = [b]_j$. Let G_j be the graph on Σ_0 in which $y \neq y'$ are adjacent iff $g(u^{j\to a}) = y$ and $g(u^{j\to b}) = y'$ for some context u and some $a \approx_j b$, and let G be the graph on Σ_0 whose edge set is the union of those of $G_1, \dots, G_m$. Our hypothesis is

(H) the graph G is connected.

It depends only on g. Call j *collision-connected* if N_j is connected; this holds, for instance, if some context u^* makes $a \mapsto g(u^{*\,j\to a})$ constant on Σ, and for $q = 2$ the two conditions coincide. For such j, G_j joins any two outputs obtained by changing x_j in a common context. So (H) holds if some collision-connected j has G_j connected. It also holds if every coordinate is collision-connected, because then two inputs z, z' are joined by a path that changes one coordinate at a time, and consecutive outputs on it are equal or adjacent in G. If N_j has no edges, neither has G_j. For $\kappa \in (0, 1]$ let $\Gamma_k(\kappa)$ and $\ell_\alpha = \frac{16}{\alpha}\ln\frac{1}{\alpha}$ be as in Section 8.8, and put

$$N = \frac{q(q-1)}{2}, \qquad L = m(q-1)N, \qquad \Gamma = \Gamma_{m-1}(\alpha),$$
$$M = \frac{2}{\alpha^2}\big(L\ell_\alpha\Gamma + 2\Gamma + 1\big), \qquad c_{q,m,\alpha} = \frac{\alpha}{8\,M^{qm+1}}.$$

Theorem 8.45 (Larger alphabets)**.** *In the setting above, suppose that (H) holds. Then there is $C = C(g, \mu)$ such that for every $n \ge 1$ and every tuple $\mathbf{f}$ with $\mathrm{err}_g^{\mu}(\mathbf{f}) = \delta$ there is a multi-polymorphism $\mathbf{F}$ of g with $\Pr_{\mu_0^{\otimes n}}[f_0 \neq F_0] \le C\delta^c$ and $\Pr_{\mu_j^{\otimes n}}[f_j \neq F_j] \le C\delta^c$ for all $j \in [m]$, where $c = c_{q,m,\alpha}$.*

Remark 8.46 (Examples). (a) *Semilattices.* Let $(\Sigma, \vee)$ be a finite semilattice (an associative, commutative, idempotent operation) and $g(x) = x_1 \vee \cdots \vee x_m$; for instance min or max on a chain. The join $\top$ of all elements is absorbing, so a context containing $\top$ makes g constant in x_j, and every j is collision-connected. Moreover $\Sigma_0 = \Sigma$, and for $y \in \Sigma$ the context with all entries y gives $g(u^{j\to y}) = y$ and $g(u^{j\to\top}) = \top$; so every $y \neq \top$ is adjacent to $\top$ in G_j, G_j is connected, and (H) holds. (b) *Monoids with zero.* Let $(\Sigma, \cdot)$ be a finite monoid with identity 1 and a zero element 0 ($0x = x0 = 0$), and $g(x) = x_1 \cdots x_m$; for instance multiplication modulo q, or multiplication of $k \times k$ matrices over a finite field. A context containing 0 makes g constant in x_j, so j is collision-connected, and the context with all entries 1 gives $g(u^{j\to a}) = a$, so $\Sigma_0 = \Sigma$, G_j is complete, and (H) holds. (c) *Near-unanimity operations.* If $m \ge 3$ and $g(b, a, \dots, a) = g(a, b, a, \dots, a) = \cdots = g(a, \dots, a, b) = a$ for all a, b (for instance the majority operation on Σ^3 that returns x_1 when all three entries differ), then the context with all entries a makes every coordinate insensitive, so every coordinate is

collision-connected, and (H) holds. (d) *The Boolean case.* For $q = 2$, (H) holds iff g is constant or not affine. Indeed, if g is not affine, apply Lemma 8.4 to the restriction of g to the coordinates on which it depends: this gives a collision-connected coordinate j^* on which g depends, so G_{j^*} has the edge $\{0,1\}$. If g is affine and not constant, then for each j either g does not depend on x_j, and G_j has no edge, or $g(u^{j\to 0}) \neq g(u^{j\to 1})$ for every context u, so N_j and G_j have no edge; hence G has no edge, and $|\Sigma_0| = 2$. So for $q = 2$ Theorem 8.45 reproves the first part of Theorem 8.33 for non-affine g, with a smaller exponent. (e) *Rounded mean.* Let $\Sigma = \{0, \dots, q-1\}$, $q \geq 3$, and $g(x_1, x_2) = \lfloor (x_1 + x_2)/2 \rfloor$. With the other entry equal to y, the map $a \mapsto \lfloor (a+y)/2 \rfloor$ is non-decreasing with steps 0 or 1 and takes at least two values, so no context makes g constant in x_j. But $\lfloor (a+y)/2 \rfloor = \lfloor (a+1+y)/2 \rfloor$ whenever $a + y$ is even, so the contexts $y \in \{0,1\}$ put the path $0, 1, \dots, q-1$ into N_j, and both coordinates are collision-connected. With $y = 0$ and $y = q-1$, g runs through $0, \dots, \lfloor \frac{q-1}{2} \rfloor$ and $\lfloor \frac{q-1}{2} \rfloor, \dots, q-1$ in steps of 0 or 1, so G_j contains the path $0, 1, \dots, q-1$, and (H) holds. (f) *Not covered.* If $g(u, \cdot)$ is injective for every context u and coordinate (for instance a group or quasigroup operation), then every N_j and every G_j has no edges, and (H) fails. This includes the larger-alphabet analogues of the affine functions, for which the Boolean case is [2, Theorem 1.7]. Of the 19632 functions $g\colon \{0,1,2\}^2 \to \{0,1,2\}$ that depend on both coordinates, 17118 satisfy (H) (script `alphabet_census.py`). Of the other 2514, the 12 Latin squares have no collisions at all; the others include affine functions of "quotient bits" such as $g(x,y) = \mathbf{1}\{x \neq 0\} \oplus \mathbf{1}\{y \neq 0\}$, for which $\mathcal{P}_1 = \mathcal{P}_2 = \{\{0\}, \{1,2\}\}$ and every G_j is edgeless. Requiring instead that the collision-connected coordinates alone connect the outputs leaves 16470 functions, and requiring a context that makes g constant in x_j for these coordinates leaves 8694. For $q = 2$ (and $m \leq 3$) (H) fails only for the affine functions, in accordance with (d).

Tools. Let Ω be a finite set, π a full-support distribution on Ω, and Q a stochastic matrix on Ω that is reversible with respect to π (i.e. $\pi(y)Q(y,y') = \pi(y')Q(y',y)$), positive semidefinite as an operator on $L^2(\pi)$, and irreducible. Then Q is self-adjoint on $L^2(\pi)$ with eigenvalues in $[0,1]$, and by the Perron–Frobenius theorem the eigenvalue 1 is simple, with the constants as eigenfunctions. Let $\lambda(Q) \in [0,1)$ be the largest eigenvalue on the orthogonal complement of the constants ($\lambda(Q) = 0$ if $|\Omega| = 1$). For $T \subseteq [n]$ and $h\colon \Omega^n \to \Lambda$, Λ a finite set with a fixed order, let the *plurality rounding* $\mathrm{pl}_T h$ be the T-junta whose value at y is the most likely value of $h(y_T, \mathbf{w})$ for $\mathbf{w} \sim \pi^{\otimes([n]\setminus T)}$, ties being broken by the order of Λ. For $q = 2$, $\mathrm{pl}_T = \mathrm{rd}_T^p$ (with ties broken towards 1). Every $\phi \in L^2(\pi^{\otimes n})$ has an orthogonal Efron–Stein decomposition $\phi = \sum_{S \subseteq [n]} \phi^{=S}$, where $\phi^{=S}$ depends only on the coordinates in S, and averaging over the coordinates in R maps ϕ to $\sum_{S \cap R = \emptyset} \phi^{=S}$.

Lemma 8.47 (Rounding and noise, finite alphabets)**.** *Let Ω, π, Q, Λ be as above, $T \subseteq [n]$, and $h\colon \Omega^n \to \Lambda$; write $h_c = \mathbf{1}\{h = c\}$. Let $\mathbf{y} \sim \pi^{\otimes n}$.*

(a) If $\mathbf{y}'$ agrees with $\mathbf{y}$ on T and $\mathbf{y}'_i$, $i \notin T$, are conditionally independent given $\mathbf{y}$ with $\Pr[\mathbf{y}'_i = b \mid \mathbf{y}] = Q(\mathbf{y}_i, b)$, then $\Pr[h(\mathbf{y}) \neq \mathrm{pl}_T h(\mathbf{y})] \leq \Pr[h(\mathbf{y}) \neq h(\mathbf{y}')]/(1 - \lambda(Q))$.

(b) If $R \subseteq [n]$ and $\mathbf{y}''$ is obtained from $\mathbf{y}$ by replacing the coordinates in R with fresh independent samples of π, then $\Pr[h(\mathbf{y}) \neq h(\mathbf{y}'')] = \sum_{c \in \Lambda} \sum_{S \cap R \neq \emptyset} \|h_c^{=S}\|_2^2$. In particular, $\Pr[h(\mathbf{y}) \neq \mathrm{pl}_T h(\mathbf{y})] \leq \sum_c \sum_{S \not\subseteq T} \|h_c^{=S}\|_2^2$.

(c) For every $c \in \Lambda$, $\Pr[\mathrm{pl}_T h(\mathbf{y}) = c] \leq |\Lambda| \Pr[h(\mathbf{y}) = c]$.

(d) Let $\mathcal{G}$ be a σ-algebra generated by a function of $\mathbf{y}$, and let $\mathrm{pl}_{\mathcal{G}} h$ take, on each atom of $\mathcal{G}$, the most likely value of h given that atom (ties broken by the order of Λ; every atom has positive probability, as π has full support). Then $\Pr[h(\mathbf{y}) \neq \mathrm{pl}_{\mathcal{G}} h(\mathbf{y})] \leq \sum_c \|h_c - \mathbb{E}[h_c \mid \mathcal{G}]\|_2^2$.

Proof. (a) Condition on $\mathbf{y}_T = t$, let $s = n - |T|$, $\phi(w) = h(t,w)$, $\phi_c = \mathbf{1}\{\phi = c\}$ and $m_c = \mathbb{E}\phi_c$, where $w \sim \pi^{\otimes s}$. If $\psi_0 = 1, \psi_1, \dots$ is an orthonormal eigenbasis of Q in $L^2(\pi)$ with eigenvalues $1 = \lambda_0 > \lambda(Q) = \lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, then the products $\psi_{a_1} \otimes \cdots \otimes \psi_{a_s}$ form an orthonormal eigenbasis of $Q^{\otimes s}$ in $L^2(\pi^{\otimes s})$ with eigenvalues $\prod_t \lambda_{a_t} \in [0,1]$, and these are at most $\lambda(Q)$ unless all $a_t = 0$. Hence, given $\mathbf{y}_T = t$,

$$\Pr[\phi(\mathbf{y}_{-T}) = \phi(\mathbf{y}'_{-T}) = c] = \langle \phi_c, Q^{\otimes s}\phi_c \rangle \leq m_c^2 + \lambda(Q)\,(m_c - m_c^2).$$

Summing over c and using $\sum_c m_c = 1$ gives $\Pr[\phi(\mathbf{y}_{-T}) \neq \phi(\mathbf{y}'_{-T})] \geq (1-\lambda(Q))(1-\sum_c m_c^2) \geq (1-\lambda(Q))(1-\max_c m_c)$, and $1 - \max_c m_c$ is the conditional probability that $h(\mathbf{y}) \neq \mathrm{pl}_T h(\mathbf{y})$. Average over t.

(b) $\mathbb{E}[h_c(\mathbf{y}'') \mid \mathbf{y}]$ is the average of h_c over the coordinates in R, so $\Pr[h(\mathbf{y}) = h(\mathbf{y}'') = c] = \mathbb{E}[h_c(\mathbf{y})h_c(\mathbf{y}'')] = \sum_{S \cap R = \emptyset} \|h_c^{=S}\|_2^2$. Since $\sum_c \|h_c\|_2^2 = \sum_c \Pr[h(\mathbf{y}) = c] = 1$, we get $\Pr[h(\mathbf{y}) \neq h(\mathbf{y}'')] = \sum_c \big(\|h_c\|_2^2 - \sum_{S \cap R = \emptyset} \|h_c^{=S}\|_2^2\big) = \sum_c \sum_{S \cap R \neq \emptyset} \|h_c^{=S}\|_2^2$. The last claim is (a) with $Q(y,b) = \pi(b)$, i.e. $\lambda(Q) = 0$ and $\mathbf{y}' = \mathbf{y}''$ for $R = [n] \setminus T$.

(c) If $\mathrm{pl}_T h(t) = c$ then $\Pr[h(\mathbf{y}) = c \mid \mathbf{y}_T = t] \geq 1/|\Lambda|$. Hence $\Pr[\mathrm{pl}_T h(\mathbf{y}) = c] \leq |\Lambda|\,\mathbb{E}\big[\mathbf{1}\{\mathrm{pl}_T h(\mathbf{y}) = c\} \Pr[h(\mathbf{y}) = c \mid \mathbf{y}_T]\big] \leq |\Lambda|\,\mathbb{E}\big[\Pr[h(\mathbf{y}) = c \mid \mathbf{y}_T]\big] = |\Lambda| \Pr[h(\mathbf{y}) = c]$.

(d) With $m_c = \Pr[h(\mathbf{y}) = c \mid \mathcal{G}]$, the conditional probability of $h \neq \mathrm{pl}_{\mathcal{G}} h$ is $1 - \max_c m_c \leq 1 - \sum_c m_c^2$. Taking expectations, $\Pr[h \neq \mathrm{pl}_{\mathcal{G}} h] \leq \sum_c \big(\|h_c\|_2^2 - \|\mathbb{E}[h_c \mid \mathcal{G}]\|_2^2\big) = \sum_c \|h_c - \mathbb{E}[h_c \mid \mathcal{G}]\|_2^2$, using $\sum_c \|h_c\|_2^2 = 1$. □

For $z \in \Sigma^m$ let $\mathbf{k}(z) = ([z_1]_1, \dots, [z_m]_m)$ be its *component vector*. For $\mathbf{z} \sim \mu$ let $\mathbf{z}'$ be a fresh sample of the conditional law of $\mathbf{z}$ given $\mathbf{k}(\mathbf{z})$ (i.e. given $\mathbf{k}(\mathbf{z}) = k$, $\mathbf{z}$ and $\mathbf{z}'$ are independent with the law of $\mathbf{z}$ conditioned on $\mathbf{k}(\mathbf{z}) = k$), and let $Q(y,y') = \Pr[g(\mathbf{z}') = y' \mid g(\mathbf{z}) = y]$ for $y, y' \in \Sigma_0$.

Lemma 8.48 (The output kernel)**.** *$\mathbf{z}' \sim \mu$, the pair $(\mathbf{z}, \mathbf{z}')$ is exchangeable, and Q is a stochastic matrix on Σ_0 that is reversible with respect to μ_0 and positive semidefinite on $L^2(\mu_0)$. If (H) holds, then Q is irreducible and $\lambda_* := \lambda(Q) < 1$.*

Proof. Given $\mathbf{k}(\mathbf{z}) = k$, the vectors $\mathbf{z}, \mathbf{z}'$ are independent and identically distributed, so $(\mathbf{z}, \mathbf{z}')$ is exchangeable and $\mathbf{z}' \sim \mu$. Hence $\mu_0(y)Q(y,y') = \Pr[g(\mathbf{z}) = y,\ g(\mathbf{z}') = y']$ is symmetric in y, y'. For $h \colon \Sigma_0 \to \mathbb{R}$ let $r(k) = \mathbb{E}[h(g(\mathbf{z})) \mid \mathbf{k}(\mathbf{z}) = k]$; by the conditional independence, $\langle h, Qh \rangle_{\mu_0} = \mathbb{E}[h(g(\mathbf{z}))h(g(\mathbf{z}'))] = \mathbb{E}[r(\mathbf{k}(\mathbf{z}))^2] \geq 0$. Finally, as μ has full support, $\Pr[g(\mathbf{z}) = y, g(\mathbf{z}') = y'] > 0$ iff $g(z) = y$ and $g(z') = y'$ for some $z, z' \in \Sigma^m$ with $\mathbf{k}(z) = \mathbf{k}(z')$. If $y \neq y'$ are adjacent in G_j, witnessed by u and $a \approx_j b$, then $z = u^{j \to a}$ and $z' = u^{j \to b}$ are such a pair. So every edge of G is an edge of the transition graph of Q, and Q is irreducible under (H). □

Edge projections. Let π be a full-support distribution on Σ. For a set $e = \{a, b\} \subseteq \Sigma$ with $a \neq b$ let p_e be the stochastic matrix with $p_e(x, \cdot) = \delta_x$ for $x \notin e$ and $p_e(x,y) = \pi(y)/\pi(e)$ for $x, y \in e$ (and $p_e(x,y) = 0$ for $x \in e$, $y \notin e$). As an operator on $L^2(\pi)$, $(p_e\phi)(x) = \sum_y p_e(x,y)\phi(y)$ is the conditional expectation onto the functions that are constant on e, i.e. the orthogonal projection onto this subspace; p_e is reversible with respect to π, $\pi p_e = \pi$, and $p_e 1 = 1$. For $R \subseteq [n]$ let P_e^R be the operator on $L^2(\pi^{\otimes n})$ that applies p_e in every coordinate in R (the tensor product of p_e on the coordinates in R and the identity on the others). For a partition $\mathcal{P}$ of Σ let $\mathrm{E}^{\mathcal{P}}$ be the conditional expectation onto the functions that are constant on every block of $\mathcal{P}$ (the orthogonal projection onto them in $L^2(\pi)$), and for $T \subseteq [n]$ let $\mathrm{E}_T^{\mathcal{P}}$ apply $\mathrm{E}^{\mathcal{P}}$ in every coordinate in $[n] \setminus T$; under $\mathbf{y} \sim \pi^{\otimes n}$ it is the conditional expectation given $\mathbf{y}_T$ and the blocks of the $\mathbf{y}_i$, $i \notin T$. For the partition into one block, $\mathrm{E}_T^{\{\Sigma\}} \phi = \sum_{S \subseteq T} \phi^{=S}$ is the average over the coordinates in $[n] \setminus T$.

Lemma 8.49 (Alternating projections)**.** *Let $\mathcal{P}$ be a partition of Σ, let $e_1, \dots, e_F$ be the edges of a forest on Σ whose components are the blocks of $\mathcal{P}$ (so $F = q - |\mathcal{P}|$), $\Pi = p_{e_1} p_{e_2} \cdots p_{e_F}$ ($\Pi = I$ if $F = 0$), and $c = \|\Pi - \mathrm{E}^{\mathcal{P}}\|_{L^2(\pi)\to L^2(\pi)}$. Then $c < 1$, and for every $T \subseteq [n]$, with $S = [n] \setminus T$, and every $\phi \in L^2(\pi^{\otimes n})$,*

$$\|\phi - \mathrm{E}_T^{\mathcal{P}}\phi\|_2 \;\le\; \frac{1}{1-c} \sum_{k=1}^{F} \big\|\phi - P_{e_k}^S \phi\big\|_2 .$$

Proof. Let K be the range of $\mathrm{E}^{\mathcal{P}}$. Each p_{e_k} fixes K (an edge lies inside a block) and, being self-adjoint, maps $K^\perp$ into itself; hence so does Π, $\Pi\mathrm{E}^{\mathcal{P}} = \mathrm{E}^{\mathcal{P}}\Pi = \mathrm{E}^{\mathcal{P}}$, and $\Pi - \mathrm{E}^{\mathcal{P}}$ vanishes on K and agrees with Π on $K^\perp$. So $c = \|\Pi|_{K^\perp}\|$ ($c = 0$ if $K^\perp = \{0\}$). If $c = 1$, there is $\phi \in K^\perp$ with $\|\phi\|_2 = 1 = \|\Pi\phi\|_2$ (the unit sphere of $K^\perp$ is compact). As the p_{e_k} are orthogonal projections, $\|p_{e_F}\phi\|_2 = \|\phi\|_2$ forces $p_{e_F}\phi = \phi$, and inductively $p_{e_k}\phi = \phi$ for every k; so ϕ is constant on every edge of the forest, hence on every block, i.e. $\phi \in K \cap K^\perp = \{0\}$, a contradiction. Thus $c < 1$.

Let $A = P_{e_1}^S \cdots P_{e_F}^S$. Decompose $L^2(\pi) = K \oplus K^\perp$ in each coordinate $i \in S$: $L^2(\pi^{\otimes n})$ is the orthogonal sum of the subspaces H_U, $U \subseteq S$, spanned by the products of $\phi_i \in K^\perp$ for $i \in U$, $\phi_i \in K$ for $i \in S \setminus U$, and arbitrary $\phi_i \in L^2(\pi)$ for $i \in T$. Each H_U is invariant under A and $\mathrm{E}_T^{\mathcal{P}}$; on H_U, A acts as the tensor product of $\Pi|_{K^\perp}$ on the coordinates in U and the identity elsewhere, whose norm is $c^{|U|}$, and $\mathrm{E}_T^{\mathcal{P}}$ acts as the identity if $U = \emptyset$ and as 0 otherwise. Hence $\|A - \mathrm{E}_T^{\mathcal{P}}\| \le \max_{\emptyset \ne U \subseteq S} c^{|U|} \le c$, and $A\mathrm{E}_T^{\mathcal{P}} = \mathrm{E}_T^{\mathcal{P}}$. Therefore

$$\phi - \mathrm{E}_T^{\mathcal{P}}\phi = (\phi - A\phi) + (A - \mathrm{E}_T^{\mathcal{P}})(\phi - \mathrm{E}_T^{\mathcal{P}}\phi), \qquad \phi - A\phi = \sum_{k=1}^{F} P_{e_1}^S \cdots P_{e_{k-1}}^S \big(\phi - P_{e_k}^S \phi\big),$$

and since the P_e^S are contractions, $\|\phi - \mathrm{E}_T^{\mathcal{P}}\phi\|_2 \le \sum_k \|\phi - P_{e_k}^S\phi\|_2 + c\,\|\phi - \mathrm{E}_T^{\mathcal{P}}\phi\|_2$. □

Lemma 8.35 holds for random variables with finitely many values, and the proof of Lemma 8.36 applies verbatim to a distribution ν on Σ^k with $\nu(z) \ge \kappa$ for all $z \in \Sigma^k$ and to sets $A_\ell \subseteq \Sigma^n$ (the marginal of ν on $k-1$ coordinates has values at least $q\kappa \ge \kappa$). We use both in this generality.

Frequent values and classes. From now on fix $\mathbf{f}$, write $\delta = \mathrm{err}_g^{\mu}(\mathbf{f})$, $\mathbf{c}_j = (f_k(\mathbf{Z}^k))_{k \ne j}$, and $\pi_k(a) = \Pr[f_k(\mathbf{Z}^k) = a]$ for $k \in [m]$ and $a \in \Sigma$. Fix $0 < \theta < \eta \le 1$ with $q\theta \le 1$ and the *gap property*

$$\text{for all } k \in [m] \text{ and } a \in \Sigma: \qquad \pi_k(a) \ge \eta \quad \text{or} \quad \pi_k(a) < \theta. \tag{10}$$

Let $V_k = \{a : \pi_k(a) \ge \eta\}$ (the *frequent* values of f_k); since $\sum_a \pi_k(a) = 1 \ge q\theta$, $V_k \ne \emptyset$, and $\Pr[f_k(\mathbf{Z}^k) \notin V_k] < q\theta$. For $j \in [m]$ let $\mathcal{V}_j = \prod_{k \ne j} V_k$ be the set of *frequent contexts*, and for $a, b \in V_j$ write $a \sim_j b$ if $g(v^{j\to a}) = g(v^{j\to b})$ for all $v \in \mathcal{V}_j$. This is an equivalence relation on V_j; let $\mathcal{C}_j$ be the set of its classes, and $\kappa_j \colon \Sigma \to \mathcal{C}_j \cup \{\perp\}$ map $a \in V_j$ to its class and $a \notin V_j$ to $\perp$. For $C \in \mathcal{C}_j$ and $v \in \mathcal{V}_j$ the value $g(v^{j\to C}) := g(v^{j\to a})$, $a \in C$, is well defined.

(F1) If $C_k \in \mathcal{C}_k$ for $k \in [m]$, then g is constant on $C_1 \times \cdots \times C_m$. Indeed, for x, x' in this product, change the coordinates of x into those of x' one at a time; at each step the current context lies in $\mathcal{V}_j$ and the changed coordinate stays in its class, so the value of g does not change.

(F2) For distinct $C, C' \in \mathcal{C}_j$ there is $v \in \mathcal{V}_j$ with $g(v^{j\to C}) \ne g(v^{j\to C'})$; and $|\mathcal{C}_j| \le q$.

(F3) For every j and $v \in \mathcal{V}_j$, $\Pr[\mathbf{c}_j = v] \ge \eta^{\Gamma}$, by Lemma 8.36 applied to the columns $\mathbf{Z}^k$, $k \ne j$, whose rows are independent samples of the marginal of μ on $\Sigma^{[m]\setminus\{j\}}$ (values $\ge \alpha$), with $\Pr[f_k(\mathbf{Z}^k) = v_k] = \pi_k(v_k) \ge \eta$.

By (F2), choose for every j and every unordered pair $\{C, C'\}$ of distinct classes in $\mathcal{C}_j$ a context $v_{CC'} \in \mathcal{V}_j$ with $g(v_{CC'}^{j\to C}) \neq g(v_{CC'}^{j\to C'})$, and let $\mathcal{N}_j$ be the set of chosen contexts; $|\mathcal{N}_j| \le N$. For $v \in \mathcal{V}_j$ let $H_v(x) = g(v^{j\to f_j(x)})$ and $h_{v,c} = \mathbf{1}\{H_v = c\}$ for $c \in \Sigma_0$ (functions on Σ^n; here j is determined by v).

For every $j \in [m]$ fix a spanning forest of N_j, i.e. a spanning tree of each component, with edges $e_{j,1}, \dots, e_{j,F_j}$, where $F_j = q - |\mathcal{P}_j| \le q-1$, and, for each edge $e = e_{j,k} = \{a, b\}$, a context u_e with $g(u_e^{j\to a}) = g(u_e^{j\to b})$. Let $c_j < 1$ be the constant of Lemma 8.49 for $\pi = \mu_j$, $\mathcal{P} = \mathcal{P}_j$ and this forest ($c_j = 0$ if $F_j = 0$), $c_* = \max_j c_j$, and write $\mathrm{E}_T^j = \mathrm{E}_T^{\mathcal{P}_j}$ on $L^2(\mu_j^{\otimes n})$. These depend only on g and μ.

Lemma 8.50 (Junta lemma, larger alphabets)**.** *Assume (10) and $\delta \le 1/2$, and let $j \in [m]$. Then there is $T_j \subseteq [n]$ with $|T_j| \le \frac{16(q-1)N}{\alpha} \ln \frac{2}{\eta^\Gamma}$ such that for every $T \supseteq T_j$ and every $v \in \mathcal{N}_j$*

$$\sum_{c\in\Sigma_0} \left\| h_{v,c} - \mathrm{E}_T^j h_{v,c} \right\|_2^2 \;\le\; \frac{8(q-1)^2(2\delta)^{\alpha/4}}{\alpha(1-c_j)^2\,\eta^\Gamma},$$

the norms being those of $L^2(\mu_j^{\otimes n})$.

Proof. We use the edge projections for $\pi = \mu_j$. If $F_j = 0$, then E_T^j is the identity and $T_j = \emptyset$ works; so let $F_j \ge 1$.

One edge. Fix $v \in \mathcal{N}_j$ and an edge $e = e_{j,k} = \{a, b\}$ with context u_e, and let $P_v = \Pr[\mathbf{c}_j = v] \ge \eta^\Gamma$ (by (F3)). We show that there is $T_{e,v} \subseteq [n]$ with

$$|T_{e,v}| \le \frac{16}{\alpha} \ln \frac{2}{P_v} \qquad \text{and} \qquad \sum_c \left\| h_{v,c} - P_e^{[n]\setminus T} h_{v,c} \right\|_2^2 \le \frac{8(2\delta)^{\alpha/4}}{\alpha P_v} \quad \text{for every } T \supseteq T_{e,v}. \tag{11}$$

This follows the proof of Lemma 8.38, with these changes. *Latent decomposition:* ν_1 is the law of the vector with context u_e and an independent μ_j-distributed entry in position j, $\sigma = \alpha/2$, and $\nu_0 = (\mu - \sigma\nu_1)/(1-\sigma) \ge \alpha/2$ pointwise; both ν_0, ν_1 have j-th marginal μ_j; ζ and $\mathbf{R} = \{i : \zeta_i = 1\}$ are as there, and given ζ the entries of $\mathbf{Z}^j$ are independent samples of μ_j. *Containment:* $\mathbf{W}$ is obtained from $\mathbf{Z}$ by replacing, independently for each $i \in \mathbf{R}$, the entry $\mathbf{Z}_{ij}$ with a sample of $p_e(\mathbf{Z}_{ij}, \cdot)$. As $\mu_j p_e = \mu_j$, given ζ the rows of $\mathbf{W}$ are independent with laws ν_{ζ_i}, so $\mathbf{W}$ has the law of $\mathbf{Z}$. A row $i \in \mathbf{R}$ has context u_e, and its j-th entry either stays or moves within e; since $g(u_e^{j\to a}) = g(u_e^{j\to b})$, $g^n(\mathbf{W}) = g^n(\mathbf{Z})$. Also $\mathbf{c}_j(\mathbf{W}) = \mathbf{c}_j(\mathbf{Z})$. On $E_v = \{\mathbf{c}_j = v,\ H_v(\mathbf{Z}^j) \neq H_v(\mathbf{W}^j)\}$ we have $\mathrm{RHS}(\mathbf{Z}) = H_v(\mathbf{Z}^j) \neq H_v(\mathbf{W}^j) = \mathrm{RHS}(\mathbf{W})$ while $f_0(g^n(\mathbf{Z})) = f_0(g^n(\mathbf{W}))$, so the test fails on $\mathbf{Z}$ or on $\mathbf{W}$, and $\Pr[E_v] \le 2\delta$. *Reverse hypercontractivity:* condition on ζ, and let $\mathbf{x}_i = (\mathbf{Z}_{ij}, \mathbf{W}_{ij})$ and $\mathbf{y}_i = (\mathbf{z}_i)_{-j}$. For $i \in \mathbf{R}$, $\mathbf{y}_i = u_e$ is constant; for $i \notin \mathbf{R}$, $\Pr[\mathbf{x}_i = (x,x), \mathbf{y}_i = u \mid \zeta] = \nu_0(u^{j\to x}) \ge \alpha/2$ and $\Pr[\mathbf{x}_i = (x,x')] = 0$ for $x \neq x'$. So Lemma 8.35 with $\kappa = \alpha/2$, $r = r_{\alpha/2}$, applied to the indicators of $\{H_v(\mathbf{Z}^j) \neq H_v(\mathbf{W}^j)\}$ and $\{\mathbf{c}_j = v\}$, gives $\Pr[E_v \mid \zeta] \ge (\Pr[H_v(\mathbf{Z}^j) \neq H_v(\mathbf{W}^j) \mid \zeta]\,\omega_v(\zeta))^{1/r}$ with $\omega_v(\zeta) = \Pr[\mathbf{c}_j = v \mid \zeta]$. *The first factor:* given ζ, $\mathbf{Z}^j \sim \mu_j^{\otimes n}$ and $\mathbb{E}[h_{v,c}(\mathbf{W}^j) \mid \mathbf{Z}^j, \zeta] = (P_e^{\mathbf{R}} h_{v,c})(\mathbf{Z}^j)$, so, as $P_e^{\mathbf{R}}$ is an orthogonal projection and $\sum_c \|h_{v,c}\|_2^2 = 1$,

$$\Pr[H_v(\mathbf{Z}^j) \neq H_v(\mathbf{W}^j) \mid \zeta] = 1 - \sum_c \langle h_{v,c}, P_e^{\mathbf{R}} h_{v,c}\rangle = \sum_c \|h_{v,c} - P_e^{\mathbf{R}} h_{v,c}\|_2^2 =: W_{e,v}(\mathbf{R}).$$

To evaluate $W_{e,v}$, let $\psi_0 = 1, \psi_1, \dots, \psi_{q-2}$ be an orthonormal basis of the range of p_e (the functions constant on e, a space of dimension $q-1$) and ψ_{q-1} a unit vector orthogonal to it, so $p_e\psi_t = \psi_t$ for $t < q-1$ and $p_e\psi_{q-1} = 0$. Expand $h_{v,c} = \sum_{\mathbf{a}\in\{0,\dots,q-1\}^n} \hat{h}_{v,c}(\mathbf{a})\,\psi_{a_1} \otimes \cdots \otimes \psi_{a_n}$ and let $D(\mathbf{a}) = \{i : a_i = q-1\}$. Then P_e^R keeps the terms with $D(\mathbf{a}) \cap R = \emptyset$ and kills the

others, so $W_{e,v}(R) = \sum_c \sum_{\mathbf{a}:\, D(\mathbf{a})\cap R\neq\emptyset} \hat{h}_{v,c}(\mathbf{a})^2$. *Jensen and the junta:* exactly as in Lemma 8.38, $\sum_c \sum_{\mathbf{a}} \hat{h}_{v,c}(\mathbf{a})^2\, \mathbb{E}[\omega_v(\zeta)\mathbf{1}\{D(\mathbf{a})\cap\mathbf{R}\neq\emptyset\}] = \mathbb{E}[W_{e,v}(\mathbf{R})\omega_v(\zeta)] \le (2\delta)^r$. Let $\varpi_i = \mathbb{E}[\omega_v(\zeta)\mathbf{1}\{i\in\mathbf{R}\}]$ and $T_{e,v} = \{i : \varpi_i \le \sigma P_v/4\}$. If $T \supseteq T_{e,v}$ and $D(\mathbf{a}) \not\subseteq T$, then $D(\mathbf{a}) \not\subseteq T_{e,v}$ and $\mathbb{E}[\omega_v(\zeta)\mathbf{1}\{D(\mathbf{a})\cap\mathbf{R}\neq\emptyset\}] > \sigma P_v/4$; hence $\sum_c \|h_{v,c} - P_e^{[n]\setminus T} h_{v,c}\|_2^2 = \sum_c \sum_{D(\mathbf{a})\not\subseteq T} \hat{h}_{v,c}(\mathbf{a})^2 \le \frac{4(2\delta)^r}{\sigma P_v}$, and $(2\delta)^r \le (2\delta)^{\alpha/4}$ gives the second part of (11). The bound on $|T_{e,v}|$ is proved word for word as part (a) of Lemma 8.38 (it uses only $\mathbb{E}[\omega_v(\zeta)] = P_v$, $0 \le \omega_v \le 1$ and the independence of the ζ_i).

All edges. For $v \in \mathcal{N}_j$ let $T_v = \bigcup_k T_{e_{j,k},v}$, so $|T_v| \le \frac{16(q-1)}{\alpha} \ln \frac{2}{\eta^\Gamma}$, and let $T_j = \bigcup_{v\in\mathcal{N}_j} T_v$, so $|T_j| \le \frac{16(q-1)N}{\alpha} \ln \frac{2}{\eta^\Gamma}$. Let $T \supseteq T_j$ and $S = [n]\setminus T$. By Lemma 8.49 (with $\pi = \mu_j$ and $\mathcal{P} = \mathcal{P}_j$), the Cauchy–Schwarz inequality, and (11) for each edge (note $T \supseteq T_{e_{j,k},v}$),

$$\sum_c \|h_{v,c} - \mathrm{E}_T^j h_{v,c}\|_2^2 \le \frac{F_j}{(1-c_j)^2} \sum_{k=1}^{F_j} \sum_c \left\|h_{v,c} - P_{e_{j,k}}^S h_{v,c}\right\|_2^2 \le \frac{8F_j^2(2\delta)^{\alpha/4}}{\alpha(1-c_j)^2 P_v},$$

and $F_j \le q-1$, $P_v \ge \eta^\Gamma$ finish the proof. □

Lemma 8.51 (From the columns to f_0, larger alphabets)**.** *Assume (10) and (H), let $T \subseteq [n]$, and suppose that $\sum_c \|h_{v,c} - \mathrm{E}_T^j h_{v,c}\|_2^2 \le \gamma$ for every $j \in [m]$ and every $v \in \mathcal{N}_j$. Then*

$$\Pr_{\mu_0^{\otimes n}}[f_0 \neq \mathrm{pl}_T f_0] \;\le\; \frac{2\delta + 2mq\theta + 2mN\gamma}{1-\lambda_*}.$$

Proof. Let $\mathbf{W}$ be obtained from $\mathbf{Z}$ by replacing, independently for each $i \notin T$, the row $\mathbf{z}_i$ with a fresh sample of the conditional law of $\mathbf{z} \sim \mu$ given $\mathbf{k}(\mathbf{z}) = \mathbf{k}(\mathbf{z}_i)$ (the pair of Lemma 8.48). The rows of $\mathbf{W}$ are independent samples of μ, so $\mathbf{W}$ has the law of $\mathbf{Z}$ and $\mathbf{W}^j \sim \mu_j^{\otimes n}$. For every j, $\mathbf{W}^j$ agrees with $\mathbf{Z}^j$ on T, and $[\mathbf{W}_{ij}]_j = [\mathbf{Z}_{ij}]_j$ for every i. For $v \in \mathcal{N}_j$ let $\phi_v = \mathrm{pl}_{\mathcal{G}_j} H_v$, where $\mathcal{G}_j$ is the σ-algebra on Σ^n generated by $x \mapsto (x_T, ([x_i]_j)_{i\notin T})$; then $\phi_v(\mathbf{W}^j) = \phi_v(\mathbf{Z}^j)$. Under $\mu_j^{\otimes n}$, $\mathbb{E}[h_{v,c} \mid \mathcal{G}_j] = \mathrm{E}_T^j h_{v,c}$, so Lemma 8.47(d) gives $\Pr[H_v(\mathbf{Z}^j) \neq \phi_v(\mathbf{Z}^j)] = \Pr[H_v(\mathbf{W}^j) \neq \phi_v(\mathbf{W}^j)] \le \gamma$. Suppose that for every j: $f_j(\mathbf{Z}^j), f_j(\mathbf{W}^j) \in V_j$, and $H_v(\mathbf{Z}^j) = \phi_v(\mathbf{Z}^j)$ and $H_v(\mathbf{W}^j) = \phi_v(\mathbf{W}^j)$ for all $v \in \mathcal{N}_j$. Then $H_v(\mathbf{Z}^j) = H_v(\mathbf{W}^j)$ for all $v \in \mathcal{N}_j$, so $f_j(\mathbf{Z}^j)$ and $f_j(\mathbf{W}^j)$ lie in the same class: two distinct classes C, C' would give $H_v(\mathbf{Z}^j) = g(v^{j\to C}) \neq g(v^{j\to C'}) = H_v(\mathbf{W}^j)$ for $v = v_{CC'}$. So the inputs $(f_k(\mathbf{Z}^k))_k$ and $(f_k(\mathbf{W}^k))_k$ of g lie in the same product of classes, and $\mathrm{RHS}(\mathbf{Z}) = \mathrm{RHS}(\mathbf{W})$ by (F1). Hence $\Pr[\mathrm{RHS}(\mathbf{Z}) \neq \mathrm{RHS}(\mathbf{W})] \le m(2q\theta + 2N\gamma)$, and therefore $\mathbf{y} = g^n(\mathbf{Z})$ and $\mathbf{y}' = g^n(\mathbf{W})$ satisfy $\Pr[f_0(\mathbf{y}) \neq f_0(\mathbf{y}')] \le 2\delta + 2mq\theta + 2mN\gamma$. Finally $\mathbf{y} \sim \mu_0^{\otimes n}$ and $\mathbf{y}'_T = \mathbf{y}_T$; the pairs $(\mathbf{y}_i, \mathbf{y}'_i)$, $i \notin T$, are independent of each other and of $\mathbf{y}_T$, each with the law $\Pr[\mathbf{y}_i = y, \mathbf{y}'_i = y'] = \Pr[g(\mathbf{z}) = y,\ g(\mathbf{z}') = y'] = \mu_0(y)Q(y,y')$. Hence, given $\mathbf{y}$, the $\mathbf{y}'_i$, $i \notin T$, are independent with $\Pr[\mathbf{y}'_i = b \mid \mathbf{y}] = Q(\mathbf{y}_i, b)$. Lemma 8.47(a) with $\Omega = \Sigma_0$, $\pi = \mu_0$, the kernel Q of Lemma 8.48 (irreducible by (H)) and $h = f_0|_{\Sigma_0^n}$ finishes the proof. □

Lemma 8.52 (From f_0 to every column, larger alphabets)**.** *Assume (10), let $T \subseteq [n]$, and suppose $\Pr_{\mu_0^{\otimes n}}[f_0 \neq J_0] \le \beta$ for some T-junta J_0. Then for every $j \in [m]$ there is a T-junta $R_j \colon \Sigma^n \to \mathcal{C}_j$ with*

$$\Pr_{\mu_j^{\otimes n}}[\kappa_j \circ f_j \neq R_j] \;\le\; q^2(2\delta + 2\beta)^{\alpha^2/2}\eta^{-2\Gamma} + (q+3)q\theta.$$

Proof. Let $\mathbf{W}$ be obtained from $\mathbf{Z}$ by replacing every row $\mathbf{z}_i$, $i \notin T$, with a fresh independent sample of μ. As in the proof of Lemma 8.40, $\mathbf{W}$ has the law of $\mathbf{Z}$, $J_0(g^n(\mathbf{Z})) = J_0(g^n(\mathbf{W}))$, and $\Pr[\mathrm{RHS}(\mathbf{Z}) \neq \mathrm{RHS}(\mathbf{W})] \le 2\delta + 2\beta$. Let $H = \kappa_j \circ f_j$, with values in $\Lambda = \mathcal{C}_j \cup \{\perp\}$, $|\Lambda| \le q+1$. Fix an ordered pair (C, C') of distinct classes, let $v = v_{CC'}$ be a context given by (F2), and let

$D = \{H(\mathbf{Z}^j) = C,\ H(\mathbf{W}^j) = C'\}$ and $E = \{\mathbf{c}_j(\mathbf{Z}) = \mathbf{c}_j(\mathbf{W}) = v\}$. On $D \cap E$, $\mathrm{RHS}(\mathbf{Z}) = g(v^{j\to C}) \neq g(v^{j\to C'}) = \mathrm{RHS}(\mathbf{W})$, so $\Pr[D \cap E] \le 2\delta + 2\beta$. Exactly as in the proof of Lemma 8.40, $\Pr[E] \ge \Pr[\mathbf{c}_j = v]^2 \ge \eta^{2\Gamma}$ (Cauchy–Schwarz and (F3)), and Lemma 8.35 with $\kappa = \alpha^2$ gives $\Pr[D \cap E] \ge (\Pr[D]\Pr[E])^{2/\alpha^2}$. Hence $\Pr[D] \le (2\delta + 2\beta)^{\alpha^2/2}\eta^{-2\Gamma}$. Summing over the at most $q(q-1)$ ordered pairs and adding $\Pr[H(\mathbf{Z}^j) = \perp] + \Pr[H(\mathbf{W}^j) = \perp] < 2q\theta$,

$$\Pr[H(\mathbf{Z}^j) \neq H(\mathbf{W}^j)] \ \le\ q(q-1)(2\delta + 2\beta)^{\alpha^2/2}\eta^{-2\Gamma} + 2q\theta.$$

$\mathbf{W}^j$ agrees with $\mathbf{Z}^j$ on T and is a fresh sample of μ_j off T, so by Lemma 8.47(b) (with $\pi = \mu_j$) the same bound holds for $\Pr[H \neq \mathrm{pl}_T H]$. Let $R_j = \mathrm{pl}_T H$ where $\mathrm{pl}_T H \neq \perp$, and let R_j be the class of the least element of V_j (for a fixed order of Σ) where $\mathrm{pl}_T H = \perp$. By Lemma 8.47(c), $\Pr[\mathrm{pl}_T H = \perp] \le (q+1)\Pr[H = \perp] < (q+1)q\theta$, and $\Pr[H \neq R_j] \le \Pr[H \neq \mathrm{pl}_T H] + \Pr[\mathrm{pl}_T H = \perp]$. □

Lemma 8.53 (Exactness on a junta, larger alphabets)**.** *Let $T \subseteq [n]$, and let $\mathbf{F}$ be a tuple such that $F_0(g^n(Z))$ and $g(F_1(Z^1), \dots, F_m(Z^m))$ depend only on the rows of Z indexed by T. If $\mathrm{err}_g^{\mu}(\mathbf{F}) < \alpha^{|T|}$, then $\mathbf{F}$ is a multi-polymorphism of g.*

Proof. As in Lemma 8.42: whether the identity holds depends only on the rows in T, and every value of these rows has probability at least $\alpha^{|T|}$. □

Proof of Theorem 8.45. Write $\lambda = \lambda_* < 1$ (Lemma 8.48 and (H)), $\ell = \ell_\alpha$, $K = 8/\alpha$, $c = c_{q,m,\alpha} = 1/(KM^{qm+1})$, and

$$c_2 = \frac{2 + 2mq + 32m(q-1)^2 N/(\alpha(1-c_*)^2)}{1-\lambda},$$
$$c_4 = 1 + c_2 + m\big(q^2(2+2c_2) + q(q+3)\big), \qquad \varepsilon_0 = \frac{2^{-L\ell}}{c_4}.$$

Reductions. If $\delta = 0$, take $\mathbf{F} = \mathbf{f}$. The all-dictator tuple $F_0 = \cdots = F_m = (x \mapsto x_1)$ is a multi-polymorphism of every g, since both sides equal $g(\mathbf{z}_1)$. So it suffices to treat $0 < \delta < \delta_0 := \varepsilon_0^{1/c}$, provided $C \ge \delta_0^{-c}$. Put $\varepsilon = \delta^c < \varepsilon_0$ and $\varepsilon_i = \varepsilon^{M^{i-1}}$ for $i = 1, \dots, qm+2$; then $\delta = \varepsilon_{qm+2}^K$. Note $\varepsilon_0 \le 2^{-\ell} \le \alpha \le q^{-2}$, because $\ell \ln 2 \ge \ln \frac{1}{\alpha}$.

The gap. The qm numbers $\pi_k(a)$ cannot meet all of the $qm+1$ disjoint intervals $[\varepsilon_{i+1}, \varepsilon_i)$, $i \in [qm+1]$. Fix i such that none of them lies in $[\varepsilon_{i+1}, \varepsilon_i)$, and let $\eta = \varepsilon_i \le \varepsilon$ and $\theta = \varepsilon_{i+1} = \eta^M$. Then (10) holds, $q\theta \le 1$, and $\delta \le \theta^K \le \theta$. Since $M \ge L\ell\Gamma + 1$ and $M\alpha^2/2 - 2\Gamma = L\ell\Gamma + 1$,

$$\theta \le \eta^{L\ell\Gamma+1} \qquad \text{and} \qquad \theta^{\alpha^2/2}\eta^{-2\Gamma} = \eta^{L\ell\Gamma+1}. \tag{12}$$

Step 1: the columns. For each $j \in [m]$, Lemma 8.50 gives T_j with $|T_j| \le \frac{16(q-1)N}{\alpha}(\ln 2 + \Gamma \ln \frac{1}{\eta})$. Let $T = \bigcup_j T_j$; then $|T| \le \frac{16L}{\alpha}(\ln 2 + \Gamma \ln \frac{1}{\eta})$. As $2\delta \le 1$ and $\delta \le \theta^{8/\alpha}$, $(2\delta)^{\alpha/4} \le 2\delta^{\alpha/4} \le 2\theta^2$, and $\theta/\eta^\Gamma = \eta^{M-\Gamma} \le 1$; so, since $T \supseteq T_j$, Lemma 8.50 gives $\sum_c \|h_{v,c} - \mathrm{E}_T^j h_{v,c}\|_2^2 \le \gamma := 16(q-1)^2\theta/(\alpha(1-c_*)^2)$ for all j and $v \in \mathcal{N}_j$.

Step 2: f_0. By Lemma 8.51, $\beta_0 := \Pr_{\mu_0^{\otimes n}}[f_0 \neq \mathrm{pl}_T f_0] \le \big(2\theta + 2mq\theta + 2mN\gamma\big)/(1-\lambda) = c_2\theta$.

Step 3: all columns. By Lemma 8.52 with $J_0 = \mathrm{pl}_T f_0$ and $\beta = \beta_0$, and by (12), there are T-juntas $R_k \colon \Sigma^n \to \mathcal{C}_k$ with

$$\beta_k := \Pr[\kappa_k \circ f_k \neq R_k] \le q^2\big((2+2c_2)\theta\big)^{\alpha^2/2}\eta^{-2\Gamma} + (q+3)q\theta$$
$$\le \big(q^2(2+2c_2) + q(q+3)\big)\eta^{L\ell\Gamma+1} \qquad (k \in [m]).$$

Step 4: the exact tuple. For each non-empty set $C \subseteq \Sigma$ fix $\mathrm{rep}(C) \in C$. Let $F_k(x) = f_k(x)$ if $\kappa_k(f_k(x)) = R_k(x)$, and $F_k(x) = \mathrm{rep}(R_k(x))$ otherwise ($k \in [m]$); let $F_0 = \mathrm{pl}_T f_0$ on Σ_0^n and $F_0 = f_0$ elsewhere. Then $\Pr[f_k \neq F_k] \leq \beta_k$, $\Pr[f_0 \neq F_0] = \beta_0$, and $F_k(x) \in R_k(x)$ for all x. By (F1), $g(F_1(Z^1), \dots, F_m(Z^m))$ depends only on the classes $R_k(Z^k)$, hence only on the rows of Z in T; and $F_0(g^n(Z))$ depends only on $g^n(Z)_T$. By the union bound, $\delta \leq \theta$ and (12),

$$\mathrm{err}_g^{\mu}(\mathbf{F}) \leq \delta + \beta_0 + \sum_{k=1}^{m} \beta_k \leq c_4\, \eta^{L\ell\Gamma+1} < 2^{-L\ell} \eta^{L\ell\Gamma} \leq \alpha^{|T|},$$

where we used $\eta \leq \varepsilon < \varepsilon_0$ and $|T| \ln \frac{1}{\alpha} \leq L\ell(\ln 2 + \Gamma \ln \frac{1}{\eta})$. By Lemma 8.53, $\mathbf{F}$ is a multi-polymorphism, and all its distances to $\mathbf{f}$ are at most $c_4 \eta^{L\ell\Gamma+1} \leq \eta \leq \varepsilon = \delta^c$ (as $c_4\eta < 1$ and $L\ell\Gamma \geq 1$). So the theorem holds with $C = \delta_0^{-c}$. □

Remark 8.54. The proof never needs the Case I/Case II split of the earlier proofs: the gap property (10) replaces "far from constants" by "frequent", and the negligible values are absorbed into the error. The edges of the collision graphs enter only through Lemma 8.50: each edge of N_j gives a containment step in which the entries of column j move only within that edge, and Lemma 8.49 combines the resulting partial bounds into closeness of H_v to a function of x_T and the components of the other entries. Their components also enter Lemmas 8.48 and 8.51, through the component vector $\mathbf{k}(z)$. The connectivity in (H) is used only in Lemma 8.51, through $\lambda_* < 1$ (Lemma 8.48): resampling every row off T within its component vector keeps the classes of all columns, and moves the outputs along the edges of G. For $q = 2$ and non-constant g, (H) says that some j with $A_j \neq \emptyset$ is a coordinate on which g depends. If g is invariant under exchanging two input coordinates $k, \ell \in [m]$, $\mu_k = \mu_\ell$ and $f_k = f_\ell$, then $V_k = V_\ell$, the relations $\sim_k$ and $\sim_\ell$ coincide (the exchange maps $\mathcal{V}_k$ onto $\mathcal{V}_\ell$), so $R_k = R_\ell$ and $F_k = F_\ell$: the analogue of Remark 8.44 holds for such pairs of input coordinates. The exponent $c_{q,m,\alpha}$ is very small; we made no attempt to optimize it. We do not know whether (H) can be dropped; the first test case is a g for which every $g(u, \cdot)$ is a bijection, such as addition in $\mathbb{Z}_q$, $q \geq 3$, where every N_j is edgeless; and the affine functions of quotient bits of Remark 8.46(f) suggest combining the present method with [2, Theorem 1.7]. The script `alphabet_check.py` (directory `numerics`) checks Lemmas 8.47 and 8.48, Lemma 8.49 ($c_j < 1$, the norm of $A - \mathrm{E}_T^{\mathcal{P}}$, and the inequality, for $q \in \{3, 4\}$, trees and forests), the main inequalities in the proof of Lemma 8.50 (for every forest edge: containment, the conditional identity, the reverse hypercontractive bound, the Jensen step and (11); then the combined bound), and the inequalities of Lemmas 8.51 and 8.52, by exact enumeration on small instances ($q \leq 4$; for the junta and propagation lemmas $q = 3$, $m = 2$, $n \leq 3$, g the rounded mean of Remark 8.46(e), min or multiplication mod 3).

8.10 A non-functional predicate over $\mathbb{Z}_3$

Alekseev and Filmus name

$$P_{\neq} = \{(a, b, c) \in \mathbb{Z}_3^3 : \ a + b + c \neq 0\}$$

as the simplest predicate that their techniques cannot handle [2, Section 9, first question]. No coordinate of $P_{/=}$ is flexible: for $(a, b, c) \in P_{/=}$ the value $a' = -b - c$ gives $(a', b, c) \notin P_{/=}$, and similarly for the other coordinates. A triple is in $P_{/=}$ iff exactly two of its entries are equal, and $|P_{/=}| = 18$. Here we treat $P_{/=}$ under the uniform distribution π on $P_{/=}$. The rows of $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \in (\mathbb{Z}_3^n)^3$ are independent samples of π, and $\mathbf{x}, \mathbf{y}, \mathbf{z}$ are the columns. For $f_1, f_2, f_3 \colon \mathbb{Z}_3^n \to \mathbb{Z}_3$ let

$$\mathrm{err}(\mathbf{f}) = \Pr\left[f_1(\mathbf{x}) + f_2(\mathbf{y}) + f_3(\mathbf{z}) = 0\right],$$

so $\mathbf{f}$ is a (π,δ)-approximate generalized polymorphism of $P_{/=}$ in the sense of [2] iff $\mathrm{err}(\mathbf{f}) \le \delta$. The marginals of π are uniform; distances are measured under the uniform measure on $\mathbb{Z}_3^n$, and $\mathrm{dist}(\mathbf{f},\mathbf{g}) = \max_j \Pr[f_j \ne g_j]$. A triple $\mathbf{g}$ is an (exact) generalized polymorphism iff $\mathrm{err}(\mathbf{g}) = 0$, i.e. $(g_1(x), g_2(y), g_3(z)) \in P_{/=}$ whenever every row of (x,y,z) is in $P_{/=}$. The following triples are generalized polymorphisms:

(A) for some j, the other two functions are constants u, v, and g_j never takes the value $-u-v$;

(B) $g_j(w) = \kappa w_i + d_j$ $(j = 1,2,3)$ for some $i \in [n]$, $\kappa \in \{1,2\}$ and d_1, d_2, d_3 with $d_1+d_2+d_3 = 0$.

For (B), $g_1(x) + g_2(y) + g_3(z) = \kappa(x_i + y_i + z_i) \ne 0$.

Theorem 8.55 (The predicate $a+b+c \ne 0$)**.** *(a) Every generalized polymorphism of $P_{/=}$ is of type (A) or (B).*

(b) There is an absolute constant C such that for every n and every $\mathbf{f}$ with $\mathrm{err}(\mathbf{f}) = \delta$ there is a generalized polymorphism $\mathbf{g}$ of type (A) or (B) with $\mathrm{dist}(\mathbf{f},\mathbf{g}) \le C\sqrt{\delta}$. One can take $C = 4 \cdot 10^7$; moreover $\mathrm{dist}(\mathbf{f},\mathbf{g}) \le 3\sqrt{\delta}$ if $\delta \le 10^{-15}$, and $\mathrm{dist}(\mathbf{f},\mathbf{g}) \le (1 + 56\,\delta^{1/8})\sqrt{\delta}$ if $\delta \le 10^{-16}$.

(c) The exponent $\frac{1}{2}$ and the constant 1 in (b) are optimal: if $E \subseteq \mathbb{Z}_3^n$ has $\Pr[E] = \eta \le \frac{1}{3}$, $f_1 \equiv 0$, $f_2 = \mathbf{1}_E$ and $f_3 = 1 + \mathbf{1}_E$, then $\mathrm{err}(\mathbf{f}) = \eta^2$ and $\mathrm{dist}(\mathbf{f},\mathbf{g}) \ge \eta$ for every generalized polymorphism $\mathbf{g}$.

(d) Near (B) the dependence is linear: if $\delta \le 8 \cdot 10^{-5}$ and $\mathrm{dist}(\mathbf{f},\mathbf{g}) \le \frac{1}{10}$ for some $\mathbf{g}$ of type (B), then $\sum_j \Pr[f_j \ne g'_j] \le 3\delta$ for some $\mathbf{g}'$ of type (B).

So, under the uniform distribution, the question of [2] has a positive answer for $P_{/=}$, with polynomial dependence, the optimal exponent and the sharp constant: the supremum of the possible distances at error at most δ (over all n) is $(1+o(1))\sqrt{\delta}$ as $\delta \to 0$, since η in (c) can be any number in $(0, \frac{1}{3}] \cap 3^{-n}\mathbb{Z}$. The constant 1 is only asymptotic: for fixed small n the ratio $\mathrm{dist}/\sqrt{\delta}$ can be larger (for $n = 1$ it reaches $\frac{2}{\sqrt{3}}$, as computed by an independent referee). As for AND_2 (Theorem 1.2), the dependence is not linear. For $n = 1$, (a) says that the 198 generalized polymorphisms are the 180 triples of type (A) and the 18 of type (B). Part (a) can also be obtained from a theorem of Filmus [10, Theorem 1.3]: for a predicate with full projections, if every generalized polymorphism with $n = 2$ is of *dictator type* (the g_j are functions of one common coordinate) or of *box type* $(g_1(\mathbb{Z}_3^n) \times g_2(\mathbb{Z}_3^n) \times g_3(\mathbb{Z}_3^n) \subseteq P)$, then so is every generalized polymorphism, for every n. For $P_{/=}$ the box-type triples are exactly those of type (A), since by the Cauchy–Davenport theorem two images with at least two elements each have sumset $\mathbb{Z}_3$; the dictator-type triples are of type (A) or (B) by Lemma 8.57 below; and an exhaustive search for $n = 2$ (`numerics/neq3_n2_check.py`: 13824 generalized polymorphisms, 13788 of type (A) and 36 of type (B)) verifies the hypothesis. In our proof, (a) is the case $\delta = 0$ of (b), whose approximating triple is always of type (A) or (B).

Let $\omega = e^{2\pi i/3}$. For $H\colon \mathbb{Z}_3^n \to \mathbb{C}$ and $\alpha \in \mathbb{Z}_3^n$ let $\chi_\alpha(w) = \omega^{\langle \alpha, w\rangle}$, $\hat{H}(\alpha) = \mathbb{E}[H\overline{\chi_\alpha}]$ (uniform w), $|\alpha| = |\{i : \alpha_i \ne 0\}|$ and $\mathrm{Inf}_k(H) = \sum_{\alpha:\alpha_k \ne 0} |\hat{H}(\alpha)|^2$. Then $H = \sum_\alpha \hat{H}(\alpha)\chi_\alpha$, $\sum_\alpha |\hat{H}(\alpha)|^2 = \mathbb{E}|H|^2$, and $\mathrm{Inf}_k(H)$ is the expectation, over the coordinates other than k, of the variance of H in the coordinate k. For a $\mathbb{Z}_3$-valued function h put $\mathrm{dc}(h) = 1 - \max_v \Pr[h = v]$, the distance of h to the constants. If $p_v = \Pr[h = v]$, then $1 - |\mathbb{E}\omega^h|^2 = 3(p_0p_1 + p_0p_2 + p_1p_2) \ge 3\max_v p_v\,\mathrm{dc}(h) \ge \mathrm{dc}(h)$.

Lemma 8.56. *(a) Each pair of columns is uniformly distributed on $(\mathbb{Z}_3^n)^2$.*

(b) For all $H_1, H_2, H_3\colon \mathbb{Z}_3^n \to \mathbb{C}$, $\mathbb{E}[H_1(\mathbf{x})H_2(\mathbf{y})H_3(\mathbf{z})] = \sum_\alpha \hat{H}_1(\alpha)\hat{H}_2(\alpha)\hat{H}_3(\alpha)(-\frac{1}{2})^{|\alpha|}$.

(c) With $F_j = \omega^{f_j}$, $\mathrm{err}(\mathbf{f}) = \frac{1}{3} + \frac{2}{3}\operatorname{Re}\mathbb{E}[F_1(\mathbf{x})F_2(\mathbf{y})F_3(\mathbf{z})]$.

Proof. (a) Each $(a,b) \in \mathbb{Z}_3^2$ has exactly two completions c with $(a,b,c) \in P_{/=}$, and similarly for the other pairs. (b) For $(a,b,c) \in \mathbb{Z}_3^3$,

$$\mathbb{E}_{(u,v,w)\sim\pi}\left[\omega^{au+bv+cw}\right] = \frac{1}{18}\Big(\sum_{(u,v,w)\in\mathbb{Z}_3^3} - \sum_{u+v+w=0}\Big)\omega^{au+bv+cw}$$
$$= \frac{27\cdot\mathbf{1}\{a=b=c=0\} - 9\cdot\mathbf{1}\{a=b=c\}}{18},$$

which is 1 if $a=b=c=0$, $-\frac{1}{2}$ if $a=b=c\neq 0$, and 0 otherwise. As the rows are independent, $\mathbb{E}[\chi_\alpha(\mathbf{x})\chi_\beta(\mathbf{y})\chi_\gamma(\mathbf{z})] = \mathbf{1}\{\alpha=\beta=\gamma\}(-\frac{1}{2})^{|\alpha|}$; now expand each H_j in characters. (c) For $s \in \mathbb{Z}_3$, $\mathbf{1}\{s=0\} = \frac{1}{3}(1+\omega^s+\omega^{2s})$, and $\omega^{2s} = \overline{\omega^s}$. □

Lemma 8.57 ($n=1$)**.** *For $n=1$, every generalized polymorphism of $P_{/=}$ is of type (A) or (B).*

Proof. Let (g_1,g_2,g_3) be one. Note $(a,b,b) \in P_{/=}$ iff $a \neq b$, and for $b \neq c$, $(a,b,c) \in P_{/=}$ iff $a \in \{b,c\}$, as $-b-c$ is the third element of $\mathbb{Z}_3$.

Suppose first that some g_j, say g_1, is a bijection. For each b, the two values $g_1(a)$, $a \neq b$, differ from $-g_2(b) - g_3(b)$, so $g_1(b) = -g_2(b) - g_3(b)$. For $b \neq c$, the rows (b,b,c) and (c,b,c) give $g_1(b)+g_2(b)+g_3(c) \neq 0$ and $g_1(c)+g_2(b)+g_3(c) \neq 0$, i.e. $g_3(c) \neq g_3(b)$ and $g_2(b) \neq g_2(c)$. So all three are bijections of $\mathbb{Z}_3$, hence affine: $g_j(a) = \kappa_j a + d_j$ with $\kappa_j \in \{1,2\}$. The identity $g_1(b)+g_2(b)+g_3(b) = 0$ for all b gives $\kappa_1+\kappa_2+\kappa_3 = 0$ and $d_1+d_2+d_3 = 0$, and the first forces $\kappa_1 = \kappa_2 = \kappa_3$. This is type (B).

Otherwise no g_j is a bijection. Suppose that two of them, say g_1, g_2, are not constant. Then g_1 takes one value only at one point a_0, and one other value at the two other points; similarly g_2 takes one of its values only at one point b_0. Fix c and $t = -g_3(c)$; then $g_1(a)+g_2(b) \neq t$ whenever $a+b \neq -c$. By the Cauchy–Davenport theorem the two-element images satisfy $g_1(\mathbb{Z}_3)+g_2(\mathbb{Z}_3) = \mathbb{Z}_3$, so $s_1+s_2 = t$ for some s_1, s_2 in the images, and every $(a,b) \in g_1^{-1}(s_1)\times g_2^{-1}(s_2)$ satisfies $a+b=-c$. If $g_1^{-1}(s_1)$ had two elements $a \neq a'$, then $a+b = a'+b$ for any $b \in g_2^{-1}(s_2)$, which is impossible; so $g_1^{-1}(s_1) = \{a_0\}$, similarly $g_2^{-1}(s_2) = \{b_0\}$, and $c = -a_0 - b_0$. As c was arbitrary, this is a contradiction. Hence at least two of the functions are constants, u and v say, and since every value of the remaining coordinate occurs in a row of $P_{/=}$, the third function never takes the value $-u-v$. This is type (A). □

Proof of Theorem 8.55. Every permutation of the three coordinates preserves $P_{/=}$ and π, so we may relabel f_1, f_2, f_3. For $k \in [n]$ write $w = (w_k, w')$ with $w' \in \mathbb{Z}_3^{[n]\setminus\{k\}}$, and for a function h on $\mathbb{Z}_3^n$ let $h^{w'} = h(\cdot, w') \colon \mathbb{Z}_3 \to \mathbb{Z}_3$. Let $\mathbf{x}', \mathbf{y}', \mathbf{z}'$ be the columns without row k; they are the columns of $n-1$ independent samples of π, independent of row k, and Lemma 8.56(a) applies to them. The *restriction* of $\mathbf{f}$ at k is the triple $R_k = (f_1^{\mathbf{x}'}, f_2^{\mathbf{y}'}, f_3^{\mathbf{z}'})$ of functions of one variable. Let $\delta = \mathrm{err}(\mathbf{f})$ and $\delta_j = \mathrm{dc}(f_j)$.

Step 1: restrictions. If R_k is not a generalized polymorphism, then given $\mathbf{x}', \mathbf{y}', \mathbf{z}'$ at least one of the 18 equally likely values of row k is a failure. Hence $\Pr[R_k \text{ is not a generalized polymorphism}] \leq 18\delta$ for every k.

Step 2: a dichotomy for each coordinate. Let $\beta_j(k) = \Pr[f_j^{\mathbf{w}'} \text{ is a bijection}]$ for uniform $\mathbf{w}'$. By Lemma 8.57, in a generalized polymorphism of one variable either all three functions are bijections (type (B)) or none is (in (A), two are constant and the third misses a value). Hence, for $j \neq l$, the independence of two columns gives $\beta_j(k)(1-\beta_l(k)) \leq 18\delta$. Let $\delta < 1/72$. If $\beta_j(k) \geq \frac{1}{2}$ for some j,

then $1-\beta_l(k) \le 36\delta$ for $l \ne j$, and then also $1-\beta_j(k) \le 36\delta$. Otherwise $\beta_j(k) \le 36\delta$ for all j. Call k a *B-coordinate* in the first case.

Step 3: a B-coordinate gives type (B). Let $\delta < 1/2394$, let k be a B-coordinate and $\mathrm{Bij}_j = \{w' : f_j^{w'} \text{ is a bijection}\}$. The bijections of $\mathbb{Z}_3$ are the maps $a \mapsto \kappa a + d$, $\kappa \in \{1,2\}$; write $f_j^{w'}(a) = \kappa_j(w')a + d_j(w')$ for $w' \in \mathrm{Bij}_j$, and put $\kappa_j(w') = 1$, $d_j(w') = 0$ for $w' \notin \mathrm{Bij}_j$. Let $\mathcal{G}$ be the event that $\mathbf{x}' \in \mathrm{Bij}_1$, $\mathbf{y}' \in \mathrm{Bij}_2$, $\mathbf{z}' \in \mathrm{Bij}_3$ and R_k is a generalized polymorphism; $\Pr[\mathcal{G}^c] \le 3 \cdot 36\delta + 18\delta = 126\delta$. On $\mathcal{G}$, Lemma 8.57 gives $\kappa_1(\mathbf{x}') = \kappa_2(\mathbf{y}') = \kappa_3(\mathbf{z}')$ and $d_1(\mathbf{x}') + d_2(\mathbf{y}') + d_3(\mathbf{z}') = 0$. Let $p_j(\kappa) = \Pr[\mathbf{w}' \in \mathrm{Bij}_j,\ \kappa_j(\mathbf{w}') = \kappa]$. For $j \ne l$, by independence of the j-th and l-th columns, $\sum_\kappa p_j(\kappa)p_l(\kappa)$ is the probability that both columns lie in the corresponding sets Bij with equal values of κ_j, κ_l, which is at least $\Pr[\mathcal{G}] \ge 1 - 126\delta$; and $\sum_\kappa p_j(\kappa)p_l(\kappa) \le \max_\kappa p_l(\kappa)$. So each p_l has an atom of mass at least $1-126\delta$, and these atoms are a common κ (otherwise $\sum_\kappa p_j p_l \le 252\delta < 1 - 126\delta$). Next let $D_j = \omega^{d_j}$ on $\mathbb{Z}_3^{[n]\setminus\{k\}}$, $s_j = |\hat{D}_j(0)|$ and $\sigma_j^2 = 1 - s_j^2 = \sum_{\alpha\ne 0} |\hat{D}_j(\alpha)|^2$. With $\xi = 126\delta$, $\Pr[d_1(\mathbf{x}') + d_2(\mathbf{y}') + d_3(\mathbf{z}') \ne 0] \le \xi$, and $\operatorname{Re}\omega^s = 1 - \frac{3}{2}\mathbf{1}\{s \ne 0\}$, so Lemma 8.56(b), $|\hat{D}_3| \le 1$ and the Cauchy–Schwarz inequality give

$$1 - \tfrac{3}{2}\xi \le \operatorname{Re}\mathbb{E}[D_1(\mathbf{x}')D_2(\mathbf{y}')D_3(\mathbf{z}')] \le s_1 s_2 s_3 + \tfrac{1}{2}\sigma_1\sigma_2 \le s_1 s_2 + \tfrac{1}{2}\sigma_1\sigma_2 \le 1 - \tfrac{1}{2}\sigma_1\sigma_2,$$

using $s_1 s_2 + \sigma_1\sigma_2 \le 1$. Hence $\sigma_1\sigma_2 \le 3\xi$ and $s_1 s_2 s_3 \ge 1 - 3\xi$, so $s_j \ge 1 - 3\xi$ and $\mathrm{dc}(d_j) \le \sigma_j^2 \le 6\xi = 756\delta$ for each j (the pair $1,2$ was arbitrary). Let d_j^0 be the most likely value of d_j. Since $\Pr[\sum_j d_j(\cdot) = 0,\ d_j(\cdot) = d_j^0\ \forall j] \ge 1 - 19\xi > 0$, $d_1^0 + d_2^0 + d_3^0 = 0$. Therefore $\mathbf{g} = (\kappa w_k + d_j^0)_j$ is of type (B), and

$$\Pr[f_j \ne g_j] \le \Pr[\mathbf{w}' \notin \mathrm{Bij}_j \text{ or } \kappa_j(\mathbf{w}') \ne \kappa] + \Pr[d_j(\mathbf{w}') \ne d_j^0] \le 126\delta + 756\delta = 882\delta. \tag{13}$$

Step 4: without B-coordinates, the error is close to that of independent inputs. Suppose no coordinate is a B-coordinate, so $\beta_j(k) \le 36\delta$ for all j,k. Let $\nu_j(k) = \Pr[f_j^{\mathbf{w}'} \text{ is not constant}]$. If R_k is a generalized polymorphism of type (A), at most one of its functions is non-constant, and if it is of type (B), its first function is a bijection. Hence, for $j \ne l$, the event that the j-th and the l-th functions of R_k are both non-constant has probability at most $18\delta + \beta_1(k) \le 54\delta$, and by independence of the two columns $\nu_j(k)\nu_l(k) \le 54\delta$. So for each k at most one j has $\nu_j(k) > \sqrt{54\delta}$. Let $F_j = \omega^{f_j}$ and $\rho = (54\delta)^{1/4}$. Since the variance of a function of modulus one is at most 1, $|\hat{F}_j(\alpha)|^2 \le \mathrm{Inf}_k(F_j) \le \nu_j(k)$ for every k with $\alpha_k \ne 0$. Choosing such a k for each $\alpha \ne 0$, at least two of $|\hat{F}_1(\alpha)|, |\hat{F}_2(\alpha)|, |\hat{F}_3(\alpha)|$ are at most ρ, so $|\hat{F}_1\hat{F}_2\hat{F}_3(\alpha)| \le \rho\sum_{j<l}|\hat{F}_j(\alpha)\hat{F}_l(\alpha)|$. Let $\mathbf{x}'', \mathbf{y}'', \mathbf{z}''$ be independent and uniform, and $\delta_{\mathrm{ind}} = \Pr[f_1(\mathbf{x}'') + f_2(\mathbf{y}'') + f_3(\mathbf{z}'') = 0] = \frac{1}{3} + \frac{2}{3}\operatorname{Re}\big(\hat{F}_1(0)\hat{F}_2(0)\hat{F}_3(0)\big)$. By Lemma 8.56(b),(c) and the Cauchy–Schwarz inequality,

$$|\delta - \delta_{\mathrm{ind}}| \le \frac{2}{3}\cdot\frac{1}{2}\sum_{\alpha\ne 0}|\hat{F}_1\hat{F}_2\hat{F}_3(\alpha)| \le \frac{\rho}{3}\sum_{j<l}\sum_{\alpha}|\hat{F}_j(\alpha)\hat{F}_l(\alpha)| \le \rho.$$

Step 5: two nearly constant functions. If $j \ne l$, then f_j takes two distinct values with probability at least $\delta_j/2$ each (its most likely value has probability $1 - \delta_j \ge \frac{1}{3} \ge \delta_j/2$), and similarly f_l; by the Cauchy–Davenport theorem the sums of these values cover $\mathbb{Z}_3$, so $\delta_{\mathrm{ind}} \ge \min_t \Pr[f_j(\mathbf{x}'') + f_l(\mathbf{y}'') = t] \ge \delta_j\delta_l/4$. Applied to the two largest of $\delta_1, \delta_2, \delta_3$, this shows that the two smallest are at most $\eta_0 := 2\sqrt{\delta_{\mathrm{ind}}} \le 2\sqrt{\delta + \rho}$. Relabel so that $\delta_1, \delta_2 \le \eta_0$.

Step 6: near type (A). Suppose $\delta_1, \delta_2 \le \eta_0$, and let τ be a number with $2\eta_0 \le \tau \le \frac{1}{8}$. We show that $\mathrm{dist}(\mathbf{f}, \mathbf{g}) \le \max\big(2\delta/\tau,\ \delta/(1-6\tau) + \sqrt{\delta/(1-6\tau)}\big)$ for some $\mathbf{g}$ of type (A), or $\delta \ge \frac{7}{8}\tau$. Let a, b be the most likely values of f_1, f_2, $E_1 = \{f_1 \ne a\}$, $E_2 = \{f_2 \ne b\}$, $\pi_j(t) = \Pr[f_j = a_j + t]$ ($a_1 = a$, $a_2 = b$), and $r_v = \Pr[f_3 = v]$, $r^* = r_{-a-b}$. The events $\{\mathbf{x} \notin E_1, \mathbf{y} \notin E_2, f_3(\mathbf{z}) = -a - b\}$,

$\{\mathbf{x}\in E_1, \mathbf{y}\notin E_2, f_3(\mathbf{z}) = -b - f_1(\mathbf{x})\}$ and $\{\mathbf{x}\notin E_1, \mathbf{y}\in E_2, f_3(\mathbf{z}) = -a - f_2(\mathbf{y})\}$ are disjoint failures, and by Lemma 8.56(a) their probabilities are at least $r^*(1-\delta_1-\delta_2)$, $\sum_{t\neq 0}\pi_1(t)r_{-a-b-t} - \delta_1\delta_2$ and $\sum_{t\neq 0}\pi_2(t)r_{-a-b-t} - \delta_1\delta_2$ (for instance, $\Pr[\mathbf{x}\in E_1, f_3(\mathbf{z}) = -b - f_1(\mathbf{x})] = \sum_{t\neq 0}\pi_1(t)r_{-a-b-t}$ as $\mathbf{x}, \mathbf{z}$ are independent, and $\Pr[\mathbf{x}\in E_1, \mathbf{y}\in E_2] = \delta_1\delta_2$). So, using $2\delta_1\delta_2 \le \eta_0(\delta_1+\delta_2)$,

$$\delta \ge (1-2\eta_0)r^* + \sum_{t\neq 0}\big(\pi_1(t)+\pi_2(t)\big)r_{-a-b-t} - \eta_0(\delta_1+\delta_2). \tag{14}$$

Case (α): $r_{-a-b+1}, r_{-a-b+2} \ge \tau$. Then, as $\eta_0 \le \frac{\tau}{2} \le \frac{1}{16}$, (14) gives $\delta \ge \frac{7}{8}r^* + \frac{\tau}{2}(\delta_1+\delta_2)$. Let $g_1 = a$, $g_2 = b$, and $g_3 = f_3$ except that $g_3 = -a-b+1$ where $f_3 = -a-b$; $\mathbf{g}$ is of type (A) and $\operatorname{dist}(\mathbf{f},\mathbf{g}) \le \max(\delta_1,\delta_2,r^*) \le 2\delta/\tau$. Case ($\beta$): otherwise. If $r^* \ge \tau$, then the first of the three events gives $\delta \ge r^*(1-\delta_1-\delta_2) \ge \frac{7}{8}\tau$. If $r^* < \tau$, then f_3 takes some value $c \neq -a-b$ with probability more than $1-2\tau \ge \frac{3}{4}$, so $\delta_3 < 2\tau$. Then all three functions are nearly constant, with most likely values $a_1 = a$, $a_2 = b$, $a_3 = c$ and $\sigma = a+b+c \neq 0$. Let $D_j = \{f_j = a_j - \sigma\}$ ("dangerous" values), $A_j = \{f_j = a_j + \sigma\}$ ("safe" values), $M_j = \{f_j = a_j\}$, $\pi_j^* = \Pr[D_j]$ and $\pi_j^\circ = \Pr[A_j]$, so $\delta_j = \pi_j^* + \pi_j^\circ \le h := 2\tau \le \frac{1}{4}$. The six events $\{D_1(\mathbf{x}), M_2(\mathbf{y}), M_3(\mathbf{z})\}$, $\{A_1(\mathbf{x}), A_2(\mathbf{y}), M_3(\mathbf{z})\}$ and their images under permutations of the indices are disjoint failures (the sums are $\sigma-\sigma$ and 3σ). By Lemma 8.56(a), the first has probability at least $\pi_1^*(1-\delta_2-\delta_3) \ge \pi_1^*(1-2h)$, and the second at least $\pi_1^\circ\pi_2^\circ - X - \pi_1^\circ\pi_3^*$, where $X = \Pr[A_1(\mathbf{x}), A_2(\mathbf{y}), A_3(\mathbf{z})]$. Let $u \le v \le w$ be the numbers π_j° in increasing order. By Lemma 8.56(b) with $H_j = \mathbf{1}_{A_j}$, $|\hat{\mathbf{1}}_A(\alpha)| \le \Pr[A]$ and the Cauchy–Schwarz inequality (taking for H_3 the indicator of the smallest set), $X \le uvw + \frac{1}{2}u\sqrt{vw}$. Summing the six bounds and using $uv+uw \ge 2u\sqrt{vw}$,

$$\delta \ge (1-2h)\sum_j \pi_j^* - h\sum_j \pi_j^* + uv + uw + vw - 3X \ge (1-3h)\Big(\sum_j \pi_j^* + vw\Big) \ge (1-6\tau)\Big(\sum_j \pi_j^* + v^2\Big).$$

Let j_0 be an index with $\pi_{j_0}^\circ = w$, let $g_j = a_j$ for $j \neq j_0$, and let $g_{j_0} = f_{j_0}$ except that $g_{j_0} = a_{j_0}$ on D_{j_0}. Then g_{j_0} never takes the value $a_{j_0} - \sigma = -\sum_{j\neq j_0} a_j$, so $\mathbf{g}$ is of type (A), and $\operatorname{dist}(\mathbf{f},\mathbf{g}) \le \sum_j \pi_j^* + v \le \delta/(1-6\tau) + \sqrt{\delta/(1-6\tau)}$.

Conclusion. Let $\delta \le \delta_0 := 10^{-15}$. Then $\rho \le 4.9\cdot 10^{-4}$ and $\eta_0 \le 2\sqrt{\delta+\rho} < \frac{1}{16}$. If there is a B-coordinate, (13) gives $\operatorname{dist}(\mathbf{f},\mathbf{g}) \le 882\delta \le 3\sqrt{\delta}$. Otherwise Steps 4–6 with $\tau = \frac{1}{8}$ give $\operatorname{dist}(\mathbf{f},\mathbf{g}) \le \max(16\delta, 4\delta+2\sqrt{\delta}) \le 3\sqrt{\delta}$ for some $\mathbf{g}$ of type (A), as $\delta < \frac{7}{64}$. If $\delta > \delta_0$, the triple $(0,0,1)$, which is of type (A), is within distance $1 \le \delta_0^{-1/2}\sqrt{\delta}$. This proves (b) with $C = \delta_0^{-1/2} < 4\cdot 10^7$, with $\mathbf{g}$ of type (A) or (B) in every case. If $\delta = 0$, then $\operatorname{dist}(\mathbf{f},\mathbf{g}) \le 3\sqrt{\delta} = 0$, so $\mathbf{f} = \mathbf{g}$ is of type (A) or (B); this proves (a).

For the last bound in (b), let $\delta \le 10^{-16}$ and $\tau = 4\sqrt{\delta+\rho}$. Then $\rho = (54\delta)^{1/4} \le 2.72\cdot 10^{-4}$, so $\tau \le \frac{1}{8}$; also $\tau \ge 2\eta_0$ and $\tau \ge 4\sqrt{\delta}$. In Step 6, $\frac{7}{8}\tau > \delta$, so the last alternative is impossible, and $2\delta/\tau \le \frac{1}{2}\sqrt{\delta}$. Let $x = 6\tau = 24\sqrt{\delta+\rho}$; then $x \le 0.4$, so $(1-x)^{-1/2} \le 1+x$ and $(1-x)^{-1} \le 2$, and the bound of Case (β) is at most $(1+x)\sqrt{\delta} + 2\delta$. As $\delta \le \rho$, $x \le 24\sqrt{2\rho} \le 55.9\,\delta^{1/8}$. Since 882δ and 2δ are at most $0.1\,\delta^{1/8}\sqrt{\delta}$, in all cases $\operatorname{dist}(\mathbf{f},\mathbf{g}) \le (1+56\,\delta^{1/8})\sqrt{\delta}$.

(c) A failure needs $0+\mathbf{1}_E(\mathbf{y})+1+\mathbf{1}_E(\mathbf{z}) \equiv 0 \pmod 3$, i.e. $\mathbf{y}\in E$ and $\mathbf{z}\in E$; by Lemma 8.56(a), $\operatorname{err}(\mathbf{f}) = \eta^2$. If $\mathbf{g}$ is of type (B), then g_2 takes the value 0 with probability $\frac{1}{3}$ and $f_2 = 0$ off E, so $\Pr[f_2 \neq g_2] \ge \frac{2}{3} - \eta \ge \eta$. If $\mathbf{g}$ is of type (A), two of its functions are constants, one of them is g_2 or g_3, and f_2 and f_3 differ from every constant with probability at least $\min(\eta, 1-\eta) = \eta$.

(d) If $\mathbf{g}$ is of type (B) with coordinate i and $\operatorname{dist}(\mathbf{f},\mathbf{g}) \le \frac{1}{10}$, then $f_j^{\mathbf{w}'} = g_j^{\mathbf{w}'}$ (a bijection) unless $f_j \neq g_j$ at one of the three points $(a,\mathbf{w}')$, so $\beta_j(i) \ge 1 - \frac{3}{10}$ and i is a B-coordinate. Step 3 gives $\mathbf{g}'$ of type (B) with $\delta_j' := \Pr[f_j \neq g_j'] \le 882\delta$. Let $E_j' = \{f_j \neq g_j'\}$ and $e(x) = f_1(x) - g_1'(x)$. On $\{\mathbf{x}\in E_1', \mathbf{y}\notin E_2', \mathbf{z}\notin E_3'\}$ the sum $f_1(\mathbf{x}) + f_2(\mathbf{y}) + f_3(\mathbf{z})$ equals $e(\mathbf{x}) + \kappa(\mathbf{x}_i + \mathbf{y}_i + \mathbf{z}_i)$. Given $\mathbf{x}$, the

sum $\mathbf{x}_i+\mathbf{y}_i+\mathbf{z}_i$ is uniform on $\{1,2\}$ (given $\mathbf{x}_i$, the pair $(\mathbf{y}_i,\mathbf{z}_i)$ is uniform on six pairs, three for each value of the sum), so this is a failure with conditional probability $\frac{1}{2}$. Removing $\{\mathbf{y}\in E_2'\}\cup\{\mathbf{z}\in E_3'\}$ costs at most $\delta_1'(\delta_2'+\delta_3')$, by Lemma 8.56(a). Treating the other two functions in the same way (the three events are disjoint),

$$\delta \ge \frac{1}{2}\sum_j \delta_j' - 2\sum_{j<l}\delta_j'\delta_l' \ge \Big(\frac{1}{2}-2\cdot 882\delta\Big)\sum_j \delta_j' \ge \frac{1}{3}\sum_j \delta_j'$$

(using $\sum_{j<l}\delta_j'\delta_l' \le \max_j \delta_j' \cdot \sum_j \delta_j'$) for $\delta \le 8\cdot 10^{-5}$. □

Remark 8.58. (a) The square root comes only from Case (β), three nearly constant functions whose deviations are "safe": such deviations fail only in pairs, and by pairwise independence (Lemma 8.56(a)) pairs occur with probability $\pi_j^\circ\pi_l^\circ$. The example of Theorem 8.55(c) has $\pi_j^*=0$, $u=0$ and $v=w=\eta$, so the bound $\delta \ge (1-3h)(\sum_j \pi_j^* + vw)$ of Case (β) is tight up to the factor $1-3h$. (The triple $f_1=f_2=\mathbf{1}_E$, $f_3=1+\mathbf{1}_E$, with $E=\{w : w_1=\cdots=w_k=0\}$ and $\eta=3^{-k}$, has error $3\eta^2$ and distance η; there $u=v=w=\eta$.) Near (B), and near (A) when the free function takes both allowed values with probability at least τ (Case (α), for fixed τ), the loss is linear. In (d) the constant 3 can be replaced by $\frac{2}{1-6\delta}$: for the $\mathbf{g}'$ of (d), $S=\sum_j \delta_j' \le 3\delta$, and the first inequality of the last display of the proof, which holds for every triple of type (B) in place of $\mathbf{g}'$, gives $\delta \ge \frac{S}{2} - 2\sum_{j<l}\delta_j'\delta_l' \ge \frac{S}{2} - S^2 \ge (\frac{1}{2}-3\delta)S$ (the whole display gives $\frac{2}{1-3528\delta}$). The constant 2 is optimal: changing one function of a triple $\mathbf{g}$ of type (B), say the first, on a set E of measure $\eta\le\frac{1}{3}$ gives error exactly $\frac{\eta}{2}$ (off E the sum is $\kappa(\mathbf{x}_i+\mathbf{y}_i+\mathbf{z}_i)\ne 0$, and on E a failure has conditional probability $\frac{1}{2}$, by the argument in the proof of (d) with $E_2'=E_3'=\emptyset$), while $\sum_j \Pr[f_j\ne g_j'] \ge \eta$ for every $\mathbf{g}'$ of type (B): equality holds for $\mathbf{g}'=\mathbf{g}$, and any other $\mathbf{g}'$ has some $g_j'\ne g_j$, which differs from g_j with probability at least $\frac{2}{3}$, so $\Pr[f_j\ne g_j'] \ge \frac{2}{3}-\eta \ge \eta$. For $\eta \le 1.6\cdot 10^{-4}$ this triple satisfies the hypotheses of (d). (b) The proof uses the uniform distribution through the pairwise independence of the columns and the diagonal form of Lemma 8.56(b). For other distributions, functions of two different columns can be correlated even if no coordinate is significant for both (two "majority-like" functions, say), so in Step 4 δ need not be close to δ_{ind}. Lemma 8.60 below extends Steps 1–3 and part (d) to every full-support distribution, with linear loss, and Theorem 8.61 extends (b), with a worse exponent, to every full-support distribution with two uniform marginals; there Steps 4–6 are replaced by Theorem 8.55 itself, applied to restrictions. Theorem 8.66 then covers every full-support distribution. (c) Step 4 is the point where the lack of flexibility is overcome. By Step 1 and Lemma 8.57, almost every restriction at a coordinate is a generalized polymorphism of one variable. Without B-coordinates these are of type (A), so at most one of the three functions is sensitive to that coordinate, and the three-wise Fourier terms are small: $|\delta-\delta_{\text{ind}}|\le\rho$ in Step 4. The script `neq3_check.py` (directory `numerics`) checks Lemma 8.57 by brute force, Lemma 8.56(b), and the inequalities of Steps 6 and (d) and the bound on X, by exact enumeration on random perturbations of generalized polymorphisms with $n\le 4$. It also checks Theorem 8.55(b) (largest observed ratio $\text{dist}/\sqrt{\delta}$ equal to 1), the example of Theorem 8.55(c) and the triple in (a) of this remark.

Other distributions. Alekseev and Filmus ask about every full-support distribution. Let μ be a distribution on $P_{/=}$ with full support, $\alpha=\min_{p\in P_{\ne}}\mu(p)$ (so $\alpha\le\frac{1}{18}$), and μ_1,μ_2,μ_3 its marginals. Now the rows of $(\mathbf{x},\mathbf{y},\mathbf{z})$ are independent samples of μ, $\text{err}_\mu(\mathbf{f})=\Pr[f_1(\mathbf{x})+f_2(\mathbf{y})+f_3(\mathbf{z})=0]$, distances to f_j are measured under $\mu_j^{\otimes n}$ (the law of the j-th column), and $\text{dist}(\mathbf{f},\mathbf{g})=\max_j \Pr_{\mu_j^{\otimes n}}[f_j\ne g_j]$. Since μ has full support, $\text{err}_\mu(\mathbf{g})=0$ iff $\mathbf{g}$ is a generalized polymorphism, so the exact solutions are those of Theorem 8.55(a). With $h^{w'}$ as in the proof of Theorem 8.55,

let $\beta_j(k) = \Pr[f_j^{\mathbf{w}'} \text{ is a bijection}]$ for $\mathbf{w}' \sim \mu_j^{\otimes([n]\setminus\{k\})}$. For a random variable $\mathbf{d}$ with finitely many values let $\mathrm{dc}(\mathbf{d}) = 1 - \max_v \Pr[\mathbf{d} = v]$. From here on α always denotes $\min \mu$ (and is no longer a Fourier index), and auxiliary symbols introduced inside a proof, such as σ, ρ, κ, γ, m and D_j, are local to that proof.

Lemma 8.59 (A conditional copy)**.** *Let $(\mathbf{u}, \mathbf{v})$ be a pair of random variables with finitely many values, let P be the law of $\mathbf{u}$, and let $K(a, a') = \sum_v \Pr[\mathbf{v} = v \mid \mathbf{u} = a] \Pr[\mathbf{u} = a' \mid \mathbf{v} = v]$ for a, a' in the range of $\mathbf{u}$. Suppose that $K(a, a') \geq \theta P(a')$ for all a, a', where $\theta \in (0, 1]$. Let $(\mathbf{u}_i, \mathbf{v}_i)$, $i \in [N]$, be independent copies of $(\mathbf{u}, \mathbf{v})$, put $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_N)$ and $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_N)$, and let $\mathbf{U}''$ be conditionally independent of $\mathbf{U}$ given $\mathbf{V}$, with the same conditional law. Then for every function d on the range of $\mathbf{U}$,*

$$\Pr[d(\mathbf{U}) \neq d(\mathbf{U}'')] \;\geq\; \theta \, \mathrm{dc}(d(\mathbf{U})).$$

The hypothesis holds with $\theta = \kappa$ if $\Pr[\mathbf{u} = a, \mathbf{v} = b] \geq \kappa \Pr[\mathbf{u} = a] \Pr[\mathbf{v} = b]$ for all a, b.

Proof. The pairs $(\mathbf{u}_i, \mathbf{u}_i'')$ are independent, each with law $P(a)K(a, a')$. On $L^2(P)$ the kernel K is self-adjoint and positive semidefinite, since $\langle \phi, K\psi \rangle_P = \mathbb{E}\big[\mathbb{E}[\phi(\mathbf{u}) \mid \mathbf{v}]\, \mathbb{E}[\psi(\mathbf{u}) \mid \mathbf{v}]\big]$; its second eigenvalue is the square of the maximal correlation of $(\mathbf{u}, \mathbf{v})$ [29]. Let $\Pi(a, a') = P(a')$. If $\theta = 1$, then $K = \Pi$, as both are Markov kernels and $K \geq \Pi$. If $\theta < 1$, then $K' = (K - \theta\Pi)/(1 - \theta)$ is a Markov kernel that is self-adjoint on $L^2(P)$, hence a contraction, and $\langle \phi, K\phi \rangle_P = (1 - \theta)\langle \phi, K'\phi \rangle_P \leq (1 - \theta)\|\phi\|_P^2$ whenever $\mathbb{E}_P \phi = 0$. In both cases the eigenvalues of K other than the eigenvalue 1 of the constants lie in $[0, 1 - \theta]$. The operator of $(\mathbf{U}, \mathbf{U}'')$ is $K^{\otimes N}$, whose eigenvalues on the tensor products of eigenfunctions are the products of eigenvalues of K; hence $\mathbb{E}[\phi(\mathbf{U})\phi(\mathbf{U}'')] \leq (\mathbb{E}\phi)^2 + (1 - \theta)\big(\mathbb{E}\phi^2 - (\mathbb{E}\phi)^2\big)$ for every real ϕ. With $\phi = \mathbf{1}\{d = v\}$ and $p_v = \Pr[d(\mathbf{U}) = v]$, summing over v gives $\Pr[d(\mathbf{U}) = d(\mathbf{U}'')] \leq 1 - \theta + \theta \sum_v p_v^2 \leq 1 - \theta + \theta \max_v p_v$. For the last claim, $\Pr[\mathbf{u} = a' \mid \mathbf{v} = b] \geq \kappa P(a')$ for every b, so $K(a, a') \geq \kappa P(a')$. □

Lemma 8.60 (Near the affine family, every full-support distribution)**.** *Let μ be a distribution on $P_{\not=}$ with full support, $\alpha = \min \mu$, and $\mathrm{err}_\mu(\mathbf{f}) = \delta \leq \alpha^4/16$.*

(a) If $\beta_j(k) \geq \frac{1}{2}$ for some j and k, then there is $\mathbf{g}$ of type (B) with coordinate k such that $\Pr_{\mu_l^{\otimes n}}[f_l \neq g_l] \leq 3\delta/\alpha^4$ for $l = 1, 2, 3$.

(b) In particular, this holds if $\Pr_{\mu_j^{\otimes n}}[f_j \neq g_j] \leq \alpha$ for some j and some $\mathbf{g}$ of type (B) with coordinate k.

Proof. Write $w = (w_k, w')$, let $\mathbf{x}', \mathbf{y}', \mathbf{z}'$ be the columns without row k, denote by $\mathbf{c}_l'$ the l-th of them, and let $R_k = (f_1^{\mathbf{x}'}, f_2^{\mathbf{y}'}, f_3^{\mathbf{z}'})$. The rows other than k are independent samples of μ, independent of row k. We use three facts.

(i) If R_k is not a generalized polymorphism, then some $p \in P_{\not=}$, as row k, gives a failure, and $\mu(p) \geq \alpha$; so $\Pr[R_k \text{ is not a generalized polymorphism}] \leq \delta/\alpha$.

(ii) Let $l \neq l'$. Every $(a, b) \in \mathbb{Z}_3^2$ has two completions in $P_{\not=}$, so the law $\mu_{ll'}$ of the l-th and l'-th entries of a row satisfies $\mu_{ll'}(a, b) \geq 2\alpha \geq 2\alpha\, \mu_l(a)\mu_{l'}(b)$, and Lemma 8.59 applies to these two entries with $\theta = 2\alpha$. Consequently, if $\Pr[b(\mathbf{c}_l') \neq b'(\mathbf{c}_{l'}')] \leq \varepsilon$ for some functions b, b', then $\mathrm{dc}(b(\mathbf{c}_l')) \leq \varepsilon/\alpha$: if $\mathbf{c}_l''$ is conditionally independent of $\mathbf{c}_l'$ given $\mathbf{c}_{l'}'$, with the same conditional law, then $(\mathbf{c}_l'', \mathbf{c}_{l'}')$ has the law of $(\mathbf{c}_l', \mathbf{c}_{l'}')$, so $\Pr[b(\mathbf{c}_l') \neq b(\mathbf{c}_l'')] \leq 2\varepsilon$.

(iii) For the l-th entry $\mathbf{u}$ of a row and the pair $\mathbf{v}$ of the other two entries, $K(a, a') \geq 3\alpha^2 \mu_l(a')$ in Lemma 8.59: for all a, a' there are at least three values v of $\mathbf{v}$ such that a and a' both complete v to an element of $P_{\not=}$ (if $a \neq a'$, those whose two entries sum to $-a''$, where $\{a, a', a''\} = \mathbb{Z}_3$), and each contributes $\Pr[\mathbf{v} = v \mid \mathbf{u} = a] \Pr[\mathbf{u} = a' \mid \mathbf{v} = v] \geq \alpha \cdot \alpha$.

(a) Let $\mathrm{Bij}_l = \{w' : f_l^{w'} \text{ is a bijection}\}$ and $b_l = \mathbf{1}_{\mathrm{Bij}_l}$. By Lemma 8.57, if R_k is a generalized polymorphism, then $b_1(\mathbf{x}') = b_2(\mathbf{y}') = b_3(\mathbf{z}')$. So by (i) and (ii), $\mathrm{dc}(b_l(\mathbf{c}'_l)) = \min(\beta_l(k), 1-\beta_l(k)) \le \delta/\alpha^2 < \frac{1}{2}$ for every l. As $\beta_j(k) \ge \frac{1}{2}$, this gives $1-\beta_j(k) \le \delta/\alpha^2$, and then $\beta_l(k) \ge \beta_j(k) - \Pr[b_j(\mathbf{c}'_j) \ne b_l(\mathbf{c}'_l)] \ge 1-\delta/\alpha^2-\delta/\alpha > \frac{1}{2}$, so $1-\beta_l(k) \le \delta/\alpha^2$ for every l. Define κ_l and d_l as in Step 3 of the proof of Theorem 8.55, and let $\mathcal{G}$ be the event that $\mathbf{x}' \in \mathrm{Bij}_1$, $\mathbf{y}' \in \mathrm{Bij}_2$, $\mathbf{z}' \in \mathrm{Bij}_3$ and R_k is a generalized polymorphism. Then $\xi := \Pr[\mathcal{G}^c] \le 3\delta/\alpha^2 + \delta/\alpha \le 4\delta/\alpha^2 \le \alpha^2/4$, and on $\mathcal{G}$, Lemma 8.57 gives $\kappa_1(\mathbf{x}') = \kappa_2(\mathbf{y}') = \kappa_3(\mathbf{z}')$ and $d_1(\mathbf{x}') + d_2(\mathbf{y}') + d_3(\mathbf{z}') = 0$. By (ii), $\mathrm{dc}(\kappa_l(\mathbf{c}'_l)) \le \xi/\alpha \le \frac{\alpha}{4}$ for each l; the most likely values of the κ_l coincide, since otherwise $\Pr[\kappa_l(\mathbf{c}'_l) = \kappa_{l'}(\mathbf{c}'_{l'})] \le 2\xi/\alpha < 1-\xi$ for some $l \ne l'$. Call the common value κ. Next let $\mathbf{x}''$ be conditionally independent of $\mathbf{x}'$ given $(\mathbf{y}', \mathbf{z}')$, with the same conditional law. Then $(\mathbf{x}'', \mathbf{y}', \mathbf{z}')$ has the law of $(\mathbf{x}', \mathbf{y}', \mathbf{z}')$, and on the intersection of $\mathcal{G}$ with the corresponding event for $(\mathbf{x}'', \mathbf{y}', \mathbf{z}')$ we have $d_1(\mathbf{x}') = -d_2(\mathbf{y}') - d_3(\mathbf{z}') = d_1(\mathbf{x}'')$. So $\Pr[d_1(\mathbf{x}') \ne d_1(\mathbf{x}'')] \le 2\xi$, and by (iii) and Lemma 8.59, $\mathrm{dc}(d_1(\mathbf{x}')) \le 2\xi/(3\alpha^2) \le \frac{1}{6}$; similarly for d_2 and d_3. Let d_l^0 be the most likely value of $d_l(\mathbf{c}'_l)$. As $\Pr[\sum_l d_l(\mathbf{c}'_l) = 0,\ d_l(\mathbf{c}'_l) = d_l^0\ \forall l] \ge 1-\xi-\frac{1}{2} > 0$, $d_1^0 + d_2^0 + d_3^0 = 0$. Therefore $g_l(w) = \kappa w_k + d_l^0$ defines a triple of type (B), and $f_l(w) = g_l(w)$ unless $w' \notin \mathrm{Bij}_l$, $\kappa_l(w') \ne \kappa$ or $d_l(w') \ne d_l^0$, so

$$\Pr_{\mu_l^{\otimes n}}[f_l \ne g_l] \le \frac{\delta}{\alpha^2} + \frac{4\delta}{\alpha^3} + \frac{8\delta}{3\alpha^4} \le \frac{3\delta}{\alpha^4}.$$

(b) Each $a \in \mathbb{Z}_3$ is the j-th entry of six elements of $P_{/=}$, so $\mu_j(a) \ge 6\alpha$ and $\Pr_{\mathbf{w}'}[f_j(a, \mathbf{w}') \ne g_j(a, \mathbf{w}')] \le \Pr[f_j \ne g_j]/(6\alpha) \le \frac{1}{6}$. As $g_j^{w'}$ is a bijection for every w', $f_j^{\mathbf{w}'}$ is a bijection except with probability at most $\frac{3}{6}$, so $\beta_j(k) \ge \frac{1}{2}$. □

So for every full-support μ the dependence is linear near the affine family, as in Theorem 8.55(d). Away from it we can treat the distributions with uniform marginals, i.e. $\mu_1 = \mu_2 = \mu_3$ uniform on $\mathbb{Z}_3$. These form the full-support part of an 11-dimensional affine subspace of the 17-dimensional simplex of distributions on $P_{/=}$; it contains π and every full-support distribution invariant under $(a, b, c) \mapsto (a+t, b+t, c+t)$. With a worse exponent it suffices that two of the marginals are uniform, a 13-dimensional family; Theorem 8.66 below covers every full-support distribution, with a much worse exponent. Distances are measured under $\mu_j^{\otimes n}$ as above, i.e. under the uniform measure for each uniform marginal.

Theorem 8.61 (Distributions with uniform marginals)**.** *Let μ be a distribution on $P_{/=}$ with full support, let $\alpha = \min \mu$, and let $r = r_{9\alpha}$ and $r' = r_{2\alpha}$ be as in Lemma 8.35, so $r \ge 8\alpha$ and $r' \ge \frac{16}{9}\alpha$. Let $\mathbf{f}$ be a triple of functions on $\mathbb{Z}_3^n$ with $\mathrm{err}_\mu(\mathbf{f}) = \delta$.*

(a) If μ_1, μ_2, μ_3 are uniform, there is a generalized polymorphism $\mathbf{g}$ with

$$\mathrm{dist}(\mathbf{f}, \mathbf{g}) \ \le\ 3 \cdot 10^7\, \alpha^{-3/2}\, \delta^{r/2} \ \le\ 3 \cdot 10^7\, \alpha^{-3/2}\, \delta^{4\alpha}.$$

(b) If two of the marginals, say μ_1 and μ_2, are uniform, there is a generalized polymorphism $\mathbf{g}$ with

$$\mathrm{dist}(\mathbf{f}, \mathbf{g}) \ \le\ 3 \cdot 10^7\, \alpha^{-3/2}\, \delta^{rr'/4} \ \le\ 3 \cdot 10^7\, \alpha^{-3/2}\, \delta^{32\alpha^2/9}.$$

Proof. The second inequalities hold as $r \ge 8\alpha$, $r' \ge \frac{16}{9}\alpha$ and $\delta \le 1$. In both cases μ_1 and μ_2 are uniform, and $\delta^{r/2} \le \delta^{rr'/4}$. Note $r < 9\alpha \le \frac{1}{2}$, so $\alpha^{-2r} \le \alpha^{-1}$. If $\beta_j(k) \ge \frac{1}{2}$ for some j and k, then Lemma 8.60(a) gives $\mathbf{g}$ with $\mathrm{dist}(\mathbf{f}, \mathbf{g}) \le 3\delta/\alpha^4 = 3\delta^{r/2}\delta^{1-r/2}\alpha^{-4} \le 3\alpha^{-2r}\delta^{r/2} \le 3\alpha^{-1}\delta^{r/2}$ when $\delta \le \alpha^4/16$ (as then $\delta^{1-r/2} \le \alpha^{4-2r}$), and when $\delta > \alpha^4/16$ every generalized polymorphism is within $1 < (16\delta/\alpha^4)^{r/2} = 4^r\alpha^{-2r}\delta^{r/2} \le 2\alpha^{-1}\delta^{r/2}$. So we may assume that $\beta_j(k) < \frac{1}{2}$ for all j and k.

The decomposition. Let $\sigma = 9\alpha \le \frac{1}{2}$ and $\nu = (\mu - \sigma\pi)/(1-\sigma)$. As $\pi \equiv \frac{1}{18}$ on $P_{/=}$, $\nu \ge (\alpha - \frac{\alpha}{2})/(1-\sigma) \ge \frac{\alpha}{2}$ there; ν is a distribution whose j-th marginal ν_j is uniform whenever μ_j is, since the marginals of π are uniform; and, as every pair has two completions, the law ν_{jl} of two entries of a ν-row satisfies $\nu_{jl}(a,b) \ge \alpha$, so $\nu_{jl}(a,b) \ge 9\alpha\,\nu_j(a)\nu_l(b)$ when ν_j and ν_l are uniform. Choose a random set $\mathbf{L} \subseteq [n]$ containing each i independently with probability σ, and then let the rows in $\mathbf{L}$ be independent samples of π and the rows outside $\mathbf{L}$ independent samples of ν. As $\mu = \sigma\pi + (1-\sigma)\nu$, the rows are independent samples of μ. For a fixed L, let $\rho = (\rho_1, \rho_2, \rho_3)$ be the columns of the rows outside L, so $\rho_j \in \mathbb{Z}_3^{L^c}$, and let $\mathbf{f}^{L,\rho} = (f_1(\cdot,\rho_1), f_2(\cdot,\rho_2), f_3(\cdot,\rho_3))$, a triple of functions on $\mathbb{Z}_3^L$. Given L and ρ, the rows in L are independent samples of π, so $\mathbf{f}^{L,\rho}$ is an instance of Theorem 8.55 with $|L|$ coordinates (for $L = \emptyset$ its functions are constants). Let $\delta(L,\rho)$ be its error and $\delta_L = \mathbb{E}_\rho\,\delta(L,\rho)$; then $\mathbb{E}_{\mathbf{L}}\delta_{\mathbf{L}} = \delta$. If μ_j is uniform, then ρ_j is uniform on $\mathbb{Z}_3^{L^c}$; below, dc of a function on $\mathbb{Z}_3^L$ refers to the uniform measure.

Step 1: restricted dictators are rare. For fixed L and j let $P_j(\rho_j) = i$ if $f_j(\cdot,\rho_j)$ is $\frac{1}{10}$-close to a function $\kappa w_i + d$ with $i \in L$, $\kappa \in \{1,2\}$, $d \in \mathbb{Z}_3$, and $P_j(\rho_j) = \perp$ if there is no such i; i is unique, because two distinct functions of this form differ with probability at least $\frac{2}{3}$. If $P_1(\rho_1), P_2(\rho_2), P_3(\rho_3)$ are not all equal, then $\mathbf{f}^{L,\rho}$ is not $\frac{1}{10}$-close to a generalized polymorphism: one of type (B) with coordinate i would give $P_j(\rho_j) = i$ for all j, and in one of type (A) two functions are constants and the third takes at most two values, while $\kappa w_i + d$ differs from every constant with probability $\frac{2}{3}$ and from every function with at most two values with probability at least $\frac{1}{3}$, which would give $P_j(\rho_j) = \perp$ for all j. By Theorem 8.55(b), then $\delta(L,\rho) > 10^{-15}$ (as $3\sqrt{10^{-15}} < \frac{1}{10}$). By Markov's inequality, $\Pr_\rho[P_1(\rho_1) \ne P_2(\rho_2)] \le 10^{15}\delta_L$, and by Lemma 8.59 for the first two entries of a ν-row, with $\theta = 9\alpha$, and the argument of fact (ii) in the proof of Lemma 8.60,

$$\mathrm{dc}(P_1(\rho_1)) \le \eta_L := \frac{2\cdot 10^{15}\,\delta_L}{9\alpha}.$$

Suppose $\eta_L \le \frac{1}{7}$ and the most likely value of $P_1(\rho_1)$ is some $i \in L$. Whenever $P_1(\rho_1) = i$, the function $f_1(\cdot,\rho_1)$ agrees with some $\kappa w_i + d$ outside a set of measure at most $\frac{1}{10}$, so for $\mathbf{w}''$ uniform on $\mathbb{Z}_3^{L\setminus\{i\}}$ the map $a \mapsto f_1(a,\mathbf{w}'',\rho_1)$ is a bijection except with probability at most $\frac{3}{10}$. Since $(\mathbf{w}'',\rho_1)$ is uniform on $\mathbb{Z}_3^{[n]\setminus\{i\}}$, this gives $\beta_1(i) \ge \frac{7}{10}(1-\eta_L) \ge \frac{3}{5}$, contrary to our assumption. Hence $\Pr_\rho[P_1(\rho_1) \ne \perp] \le \eta_L$ if $\eta_L \le \frac{1}{7}$, and trivially $\Pr_\rho[P_1(\rho_1) \ne \perp] \le 7\eta_L$ in all cases.

Step 2: products of distances to constants. If $P_1(\rho_1) = \perp$ and $\delta(L,\rho) \le 10^{-15}$, then Theorem 8.55(b) gives a generalized polymorphism $\mathbf{g}'$ within $\varepsilon_0 := 3\sqrt{\delta(L,\rho)} < \frac{1}{10}$ of $\mathbf{f}^{L,\rho}$. It is not of type (B), as its coordinate would be $P_1(\rho_1)$; so it is of type (A): for some labelling $\{j_1,j_2,j_3\} = [3]$, $g'_{j_1} \equiv u$, $g'_{j_2} \equiv v$ and g'_{j_3} never takes the value $-u-v$ (for $L = \emptyset$ all three functions of $\mathbf{f}^{L,\rho}$ are constants). Write $h_t = f_t(\cdot,\rho_t)$, $p_1 = \Pr[h_{j_1} \ne u] \le \varepsilon_0$ and $p_2 = \Pr[h_{j_2} \ne v] \le \varepsilon_0$; as $\mathrm{dc}(h_{j_1}) \le p_1$ and $\mathrm{dc}(h_{j_2}) \le p_2$, we have $\mathrm{dc}(h_{j_1})\,\mathrm{dc}(h_{j_2}) \le \varepsilon_0^2 = 9\,\delta(L,\rho)$. For the pairs containing j_3 we use that, given L and ρ, the rows in L are independent samples of π, so that any two of the columns restricted to L are independent and uniform (Lemma 8.56(a)). For $a \ne u$ let $q_a = \Pr[h_{j_3} = -a-v]$. The events that $h_{j_1}, h_{j_2}, h_{j_3}$ take the values $a, v, -a-v$ on the respective columns, $a \ne u$, are disjoint failures, and the probability of each is at least $\Pr[h_{j_1} = a,\ h_{j_3} = -a-v] - \Pr[h_{j_1} = a,\ h_{j_2} \ne v] = \Pr[h_{j_1} = a]\,(q_a - p_2)$. Summing over a, $\delta(L,\rho) \ge p_1\,(\min_{a\ne u} q_a - p_2)$. When a runs over $\mathbb{Z}_3 \setminus \{u\}$, $-a-v$ runs over the two values other than $-u-v$, and $\Pr[h_{j_3} = -u-v] \le \Pr[h_{j_3} \ne g'_{j_3}] \le \varepsilon_0$; so $\mathrm{dc}(h_{j_3}) \le 1 - \max_{a\ne u} q_a = \Pr[h_{j_3} = -u-v] + \min_{a\ne u} q_a \le \varepsilon_0 + \min_{a\ne u} q_a$. Hence

$$\mathrm{dc}(h_{j_1})\,\mathrm{dc}(h_{j_3}) \le p_1\big(\min_{a\ne u} q_a + \varepsilon_0\big) \le \delta(L,\rho) + p_1(p_2+\varepsilon_0) \le \delta(L,\rho) + 2\varepsilon_0^2 = 19\,\delta(L,\rho),$$

and in the same way (exchanging the roles of j_1, u and j_2, v) $\mathrm{dc}(h_{j_2})\,\mathrm{dc}(h_{j_3}) \le 19\,\delta(L,\rho)$. So $m_{jl}(L,\rho) := \mathrm{dc}(f_j(\cdot,\rho_j))\,\mathrm{dc}(f_l(\cdot,\rho_l)) \le 19\,\delta(L,\rho)$ for every pair $j \ne l$ in this case, and $m_{jl} \le 1$

always. By Step 1 and Markov's inequality,

$$\mathbb{E}_\rho\, m_{jl}(L,\rho) \le 19\,\delta_L + 7\eta_L + 10^{15}\delta_L \le K_1\delta_L, \qquad K_1 := \frac{2\cdot 10^{15}}{\alpha}.$$

Step 3: reverse hypercontractivity and Harris's inequality. Let $D_j(L) = \mathbb{E}_{\rho_j}\mathrm{dc}(f_j(\cdot,\rho_j))$. Let $j \neq l$ be such that μ_j and μ_l are uniform (any pair in (a), the pair $\{1,2\}$ in (b)). Apply Lemma 8.35 to the independent pairs $(\rho_{j,i},\rho_{l,i})$, $i \notin L$, whose law ν_{jl} satisfies the hypothesis with $\kappa = 9\alpha$, and to $\Phi(\rho_j) = \mathrm{dc}(f_j(\cdot,\rho_j))$ and $\Psi(\rho_l) = \mathrm{dc}(f_l(\cdot,\rho_l))$. These take values in $[0,1]$, $\Phi\Psi = m_{jl}$, and $\|\Phi\|_r \ge (\mathbb{E}\Phi)^{1/r}$ (as $\Phi^r \ge \Phi$); hence $(D_j(L)D_l(L))^{1/r} \le \mathbb{E}[\Phi\Psi] \le K_1\delta_L$, i.e. $D_j(L)D_l(L) \le K_1^r\delta_L^r$. Moreover D_j is nondecreasing in L: if $L \subseteq L'$, $\mathbf{s}$ is uniform on $\mathbb{Z}_3^{L'\setminus L}$ and $\rho' \in \mathbb{Z}_3^{L'^c}$, then $\max_v \Pr[f_j(\cdot,\rho') = v] = \max_v \mathbb{E}_{\mathbf{s}}\Pr[f_j(\cdot,\mathbf{s},\rho') = v] \le \mathbb{E}_{\mathbf{s}}\max_v \Pr[f_j(\cdot,\mathbf{s},\rho') = v]$, so $\mathrm{dc}(f_j(\cdot,\rho')) \ge \mathbb{E}_{\mathbf{s}}\mathrm{dc}(f_j(\cdot,\mathbf{s},\rho'))$, and $(\mathbf{s},\rho')$ is uniform on $\mathbb{Z}_3^{L^c}$ when ρ' is uniform (here we use that μ_j is uniform); the same holds for D_l. By Harris's inequality [15] for the product measure of $\mathbf{L}$ and by Jensen's inequality,

$$\mathbb{E}_{\mathbf{L}}D_j(\mathbf{L})\cdot\mathbb{E}_{\mathbf{L}}D_l(\mathbf{L}) \le \mathbb{E}_{\mathbf{L}}\big[D_j(\mathbf{L})D_l(\mathbf{L})\big] \le K_1^r\,\mathbb{E}_{\mathbf{L}}\delta_{\mathbf{L}}^r \le K_1^r\delta^r.$$

Step 4: noise. Let μ_j be uniform. If $p_0 \ge p_1, p_2$ are the probabilities of the values of h, then $1 - |\mathbb{E}\omega^h|^2 = 3(p_0p_1 + p_0p_2 + p_1p_2) \le 3(p_0+p_2)(p_1+p_2) \le 3\,\mathrm{dc}(h)$. Let $F = \omega^{f_j}$. With the notation of Lemma 8.56, $\mathbb{E}_{\rho_j}\big[1 - |\mathbb{E}_w F(w,\rho_j)|^2\big] = \sum_{\gamma:\,\gamma_i\neq 0 \text{ for some } i\in L} |\hat F(\gamma)|^2$ (the inner expectation over w uniform on $\mathbb{Z}_3^L$), so

$$\mathbb{E}_{\mathbf{L}}D_j(\mathbf{L}) \ge \frac{1}{3}\sum_{\gamma\neq 0}\big(1-(1-\sigma)^{|\gamma|}\big)|\hat F(\gamma)|^2 \ge \frac{\sigma}{3}\big(1-|\mathbb{E}\omega^{f_j}|^2\big) \ge \frac{\sigma}{3}\,\mathrm{dc}(f_j) = 3\alpha\,\mathrm{dc}(f_j).$$

Step 5: one nearly constant function gives a second (any full-support μ). Suppose $\mathrm{dc}(f_j) \le \varepsilon'$ for some j, let u be the most likely value of f_j and $\{l,m\} = [3]\setminus\{j\}$, and let $\mathbf{c}_1 = \mathbf{x}$, $\mathbf{c}_2 = \mathbf{y}$, $\mathbf{c}_3 = \mathbf{z}$. For $s \in \mathbb{Z}_3$ the events $\{f_j(\mathbf{c}_j) = u,\ f_l(\mathbf{c}_l) = s,\ f_m(\mathbf{c}_m) = -u-s\}$ are disjoint failures, so $\sum_s \Pr[f_l(\mathbf{c}_l) = s,\ f_m(\mathbf{c}_m) = -u-s] \le \delta + \varepsilon'$. The pairs $(\mathbf{c}_{l,i},\mathbf{c}_{m,i})$ are independent with law $\mu_{lm} \ge 2\alpha \ge 2\alpha\,\mu_l\mu_m$, so Lemma 8.35 with $\kappa = 2\alpha$ gives $p_sq_s \le \gamma := (\delta+\varepsilon')^{r'}$ for every s, where $p_s = \Pr[f_l(\mathbf{c}_l) = s]$ and $q_s = \Pr[f_m(\mathbf{c}_m) = -u-s]$. We claim $\min(\mathrm{dc}(f_l),\mathrm{dc}(f_m)) \le 2\sqrt{\gamma}$. This is trivial if $\sqrt{\gamma} > \frac13$, as $\mathrm{dc} \le \frac23$. Otherwise choose s^* with $q_{s^*} \ge \frac13$, so $p_{s^*} \le 3\gamma$, and suppose $\mathrm{dc}(f_l) > 2\sqrt{\gamma}$, i.e. $p_s < 1-2\sqrt{\gamma}$ for all s. Then each $s \neq s^*$ has $p_s > 1 - p_{s^*} - (1-2\sqrt{\gamma}) \ge 2\sqrt{\gamma} - 3\gamma \ge \sqrt{\gamma}$, hence $q_s \le \gamma/p_s \le \sqrt{\gamma}$, and $\mathrm{dc}(f_m) \le 1 - q_{s^*} \le 2\sqrt{\gamma}$.

Conclusion. In case (a), Steps 3 and 4 give $\mathrm{dc}(f_j)\,\mathrm{dc}(f_l) \le K_1^r\delta^r/(9\alpha^2)$ for all $j \neq l$, so two indices $j \neq l$ have $\mathrm{dc}(f_j),\mathrm{dc}(f_l) \le \varepsilon := K_1^{r/2}\delta^{r/2}/(3\alpha)$; put $\varepsilon'' = \varepsilon$. In case (b), Steps 3 and 4 give this for $\{j,l\} = \{1,2\}$, so one $j \in \{1,2\}$ has $\mathrm{dc}(f_j) \le \varepsilon$, and Step 5 with $\varepsilon' = \varepsilon$ gives a second index; so two indices have $\mathrm{dc} \le \varepsilon'' := \max\big(\varepsilon, 2(\delta+\varepsilon)^{r'/2}\big)$. Let $j \neq l$ be these indices, u, v the most likely values of f_j, f_l, and m the third index. The event that f_j, f_l, f_m take the values u, v, $-u-v$ on the respective columns is a failure, so $\Pr[f_m = -u-v] \le \delta + 2\varepsilon''$. Let $g_j \equiv u$, $g_l \equiv v$, and $g_m = f_m$ except that $g_m = 1-u-v$ where $f_m = -u-v$. Then $\mathbf{g}$ is of type (A) and $\mathrm{dist}(\mathbf{f},\mathbf{g}) \le \delta + 2\varepsilon''$. Finally $\varepsilon \le \frac{1}{3\alpha}K_1^{1/2}\delta^{r/2} = \frac{\sqrt{2}}{3}\,10^{7.5}\alpha^{-3/2}\delta^{r/2} < 1.491\cdot 10^7\alpha^{-3/2}\delta^{r/2}$, and $\delta \le \delta^{r/2} \le 18^{-3/2}\alpha^{-3/2}\delta^{r/2} < 0.014\,\alpha^{-3/2}\delta^{r/2}$, as $\alpha \le \frac{1}{18}$. This proves (a). In case (b), $\delta+\varepsilon \le 1.6\cdot 10^7\alpha^{-3/2}\delta^{r/2}$, so, as $r' \le 1$ and $\alpha^{-3/4} \le 18^{-3/4}\alpha^{-3/2}$, $2(\delta+\varepsilon)^{r'/2} \le 2\cdot 4000\,\alpha^{-3/4}\delta^{rr'/4} < 10^3\alpha^{-3/2}\delta^{rr'/4}$; hence $\varepsilon'' < 1.491\cdot 10^7\alpha^{-3/2}\delta^{rr'/4}$ and $\delta + 2\varepsilon'' < 3\cdot 10^7\alpha^{-3/2}\delta^{rr'/4}$. □

Proposition 8.62 (The exponent must depend on μ)**.** *(a) For $0 < \alpha \le \frac{1}{18}$ let $\mu^{(\alpha)}$ give mass α to each of the six elements (a,b,b) of $P_{\neq}$ and mass $\frac{1-6\alpha}{12}$ to each of the other twelve. Then $\mu^{(\alpha)}$ has full support, $\min\mu^{(\alpha)} = \alpha$ and uniform marginals, and $\mu^{(1/18)} = \pi$. For every $k \ge 1$*

there is a triple $\mathbf{f}$ *of functions on* $\mathbb{Z}_3^k$ *with* $\operatorname{err}_{\mu^{(\alpha)}}(\mathbf{f}) = (2\alpha)^k$ *and* $\operatorname{dist}(\mathbf{f},\mathbf{g}) \geq 3^{-k}$ *for every generalized polymorphism* $\mathbf{g}$*. Hence if constants* $C, c > 0$ *are such that every* $\mathbf{f}$ *(for every* n*) is within* $C \operatorname{err}_{\mu^{(\alpha)}}(\mathbf{f})^c$ *of a generalized polymorphism, then* $c \leq \log 3 / \log \frac{1}{2\alpha}$*.*

(b) For $0 < \alpha \leq \frac{1}{45}$ *let* μ_α^{sh} *give mass* $\frac{1-12\alpha}{6}$ *to each of the six elements* $(a, b, b+1)$ *of* $P_{/=}$ *and mass* α *to each of the other twelve. Then* μ_α^{sh} *has full support,* $\min \mu_\alpha^{\mathrm{sh}} = \alpha$ *and uniform marginals. If constants* $C, c > 0$ *are such that every* $\mathbf{f}$ *(for every* n*) is within* $C \operatorname{err}_{\mu_\alpha^{\mathrm{sh}}}(\mathbf{f})^c$ *of a generalized polymorphism, then* $c \leq \frac{9}{2} e^2 \alpha < 34\alpha$*.*

Proof. (a) As $\alpha \leq \frac{1}{18}$, $\frac{1-6\alpha}{12} \geq \frac{1}{18} \geq \alpha$. Each $v \in \mathbb{Z}_3$ is the first entry of two elements (v, b, b) and of four elements with distinct last two entries, and it is the second (or third) entry of two elements (a, v, v) and of four elements with distinct last two entries; so every marginal gives v probability $2\alpha + 4 \cdot \frac{1-6\alpha}{12} = \frac{1}{3}$. Let $E = \{0\} \subseteq \mathbb{Z}_3^k$, $f_1 \equiv 0$, $f_2 = \mathbf{1}_E$ and $f_3 = 1 + \mathbf{1}_E$. As in Theorem 8.55(c), a failure means $\mathbf{y} \in E$ and $\mathbf{z} \in E$, so $\operatorname{err}_{\mu^{(\alpha)}}(\mathbf{f}) = \big(\mu^{(\alpha)}(1,0,0) + \mu^{(\alpha)}(2,0,0)\big)^k = (2\alpha)^k$. The proof of Theorem 8.55(c) uses only that the second and third columns are uniform, so $\operatorname{dist}(\mathbf{f},\mathbf{g}) \geq \Pr[E] = 3^{-k}$ for every generalized polymorphism $\mathbf{g}$. If $3^{-k} \leq C(2\alpha)^{ck}$ for all k, then $3^{-1} \leq (2\alpha)^c$.

(b) The six elements of $P_{/=}$ whose third entry is the second plus 1 are the $(a, b, b+1)$ with $a \in \{b, b+1\}$, and $\frac{1-12\alpha}{6} \geq \alpha$ as $\alpha \leq \frac{1}{18}$. Adding t to all three entries maps $P_{/=}$ onto itself and preserves this condition, so it preserves μ_α^{sh}; hence every marginal is invariant under translations, i.e. uniform. Under μ_α^{sh} the third entry of a row is the second plus 1 except with probability 12α, and the triple below exploits this. Let $n = 3s$, $\tau = 2e^2\alpha$ (so $\tau < \frac{1}{3}$) and $\ell = \lceil \tau n \rceil$; for $w \in \mathbb{Z}_3^n$ let $N_v(w)$ be the number of i with $w_i = v$. Let $Y = \{w : N_0(w) \geq s + \ell\}$, $Z = \{w : N_1(w) \leq s - \ell\}$, $f_1 \equiv 0$, $f_2 = \mathbf{1}_Y$ and $f_3 = 1 + \mathbf{1}_Z$. As in Theorem 8.55(c), a failure means $\mathbf{y} \in Y$ and $\mathbf{z} \in Z$. Let $\mathbf{D}$ be the number of rows with $\mathbf{y}_i = 0$ and $\mathbf{z}_i \neq 1$. Every other row with $\mathbf{y}_i = 0$ has $\mathbf{z}_i = 1$, so $N_1(\mathbf{z}) \geq N_0(\mathbf{y}) - \mathbf{D}$, and a failure forces $\mathbf{D} \geq 2\ell$. A row has $\mathbf{y}_i = 0$ and $\mathbf{z}_i \neq 1$ with probability 4α (it is $(1,0,0)$, $(2,0,0)$, $(0,0,2)$ or $(2,0,2)$), so, by the union bound over sets of 2ℓ rows and as $2\ell \geq 2\tau n$,

$$\operatorname{err}_{\mu_\alpha^{\mathrm{sh}}}(\mathbf{f}) \leq \Pr[\mathbf{D} \geq 2\ell] \leq \binom{n}{2\ell} (4\alpha)^{2\ell} \leq \left(\frac{4e\alpha n}{2\ell}\right)^{2\ell} \leq \left(\frac{2e\alpha}{\tau}\right)^{2\ell} = e^{-2\ell} \leq e^{-2\tau n}.$$

Under the uniform measure, N_0 and N_1 have the binomial distribution $\operatorname{Bin}(n, \frac{1}{3})$, with mean s and variance $\frac{2n}{9}$, so by Chebyshev's inequality $\Pr[Y], \Pr[Z] \leq \frac{2n}{9\ell^2} \leq \frac{2}{9}$ once $n \geq \tau^{-2}$. Then $\operatorname{dist}(\mathbf{f},\mathbf{g}) \geq \min(\Pr[Y], \Pr[Z])$ for every generalized polymorphism $\mathbf{g}$, by the argument of Theorem 8.55(c): $\mathbf{g}$ is of type (A) or (B) by Theorem 8.55(a); if it is of type (B), then $\Pr[f_1 \neq g_1] = \frac{2}{3} > \frac{2}{9} \geq \min(\Pr[Y], \Pr[Z])$; if it is of type (A), then g_2 or g_3 is a constant, and f_2 and f_3 differ from every constant with probability at least $\min(\Pr[Y], 1 - \Pr[Y]) = \Pr[Y]$ and $\Pr[Z]$, respectively. Let $H(q) = -q \ln q - (1-q)\ln(1-q)$, with $0 \ln 0 = 0$, and let $d(q\|p) = q \ln \frac{q}{p} + (1-q) \ln \frac{1-q}{1-p}$ be the relative entropy between Bernoulli distributions. For $0 \leq j \leq n$, j is a most likely value of $\operatorname{Bin}(n, \frac{j}{n})$, which takes $n+1$ values, so $\binom{n}{j} \geq \frac{1}{n+1} e^{nH(j/n)}$, and hence $\Pr[\operatorname{Bin}(n,p) = j] \geq \frac{1}{n+1} e^{-nd(j/n\|p)}$. By Jensen's inequality $d(q\|p) \leq \ln\left(1 + \frac{(q-p)^2}{p(1-p)}\right) \leq \frac{(q-p)^2}{p(1-p)}$, which is $\frac{9}{2}(q - \frac{1}{3})^2$ for $p = \frac{1}{3}$. If also $n \geq (\frac{1}{3} - \tau)^{-1}$, then $0 \leq s - \ell$ and $s + \ell \leq n$, and $j = s \pm \ell$ has $|\frac{j}{n} - \frac{1}{3}| = \frac{\ell}{n} \leq \tau + \frac{1}{n}$; so

$$\Pr[Y] \geq \Pr[N_0 = s + \ell] \geq \frac{e^{-\frac{9}{2} n(\tau + 1/n)^2}}{n+1}, \qquad \Pr[Z] \geq \Pr[N_1 = s - \ell] \geq \frac{e^{-\frac{9}{2} n(\tau + 1/n)^2}}{n+1}.$$

Hence $e^{-\frac{9}{2} n(\tau+1/n)^2}/(n+1) \leq Ce^{-2c\tau n}$ for all large n divisible by 3. Taking logarithms, dividing by n and letting $n \to \infty$ gives $2c\tau \leq \frac{9}{2}\tau^2$, i.e. $c \leq \frac{9}{4}\tau = \frac{9}{2}e^2\alpha$, and $\frac{9}{2}e^2 < 33.26$. □

Remark 8.63. (a) So the question of [2] has a positive answer for $P_{/=}$, with polynomial dependence, under every full-support distribution with two uniform marginals (and, by Theorem 8.66 below, under every full-support distribution). The exponent cannot be independent of μ: by Proposition 8.62(a), for $\mu^{(\alpha)}$ with $\alpha < \frac{1}{18}$ the best exponent lies between $r_{18\alpha}/2 \geq 8\alpha$ (Proposition 8.81) and $\log 3/\log\frac{1}{2\alpha}$; both bounds tend to 0 as $\alpha \to 0$ and to $\frac{1}{2}$ as $\alpha \to \frac{1}{18}$, and for π it is $\frac{1}{2}$ (Theorem 8.55). By Proposition 8.62(b), the exponent of Theorem 8.61(a) has the right order of magnitude, and by Propositions 8.81 and 8.82 below, the supremum of the exponents c such that every full-support μ with uniform marginals and $\min\mu = \alpha < \frac{1}{18}$ admits a bound $\mathrm{dist} \leq C_\mu \mathrm{err}_\mu^c$ lies in $[r_{18\alpha}/2, 9\alpha]$, an interval whose ends are within a factor $\frac{9}{8}$ (Theorem 8.61(a) gives the exponent $r_{9\alpha}/2 \geq 4\alpha$ with the constant $3\cdot 10^7\alpha^{-3/2}$); as $9\alpha \leq \log 3/\log\frac{1}{2\alpha}$ for $\alpha \leq \frac{1}{18}$, the family $\mu^{(\alpha)}$ does not lower the upper end. The witness for μ_α^{sh} is the standard sharpness example for reverse hypercontractivity (compare [21, §3.4]): the third entry of a row is nearly a function of the second ($\mathbf{z}_i = \mathbf{y}_i + 1$ except with probability 12α), and the tail sets Y and Z give $\Pr[\mathbf{y}\in Y, \mathbf{z}\in Z] \leq e^{-2\tau n} \leq (\Pr[Y]\Pr[Z])^{(1-o(1))/(9e^2\alpha)}$ as $n\to\infty$, by the displays in the proof. So the order-α loss in the exponent of Theorem 8.61(a), which comes from reverse hypercontractivity (Step 3), cannot be avoided. As μ_α^{sh} has uniform marginals, the exponents of Theorem 8.61(b) and of Theorem 8.66 below can be improved at most to 9α (Proposition 8.82). (b) The decomposition supplies, inside μ, independent rows of π, on which the pairwise independence of Lemma 8.56(a) is available; Theorem 8.55 handles the restrictions, the pairwise independence turns its square-root bound into the linear bound on products of Step 2, and reverse hypercontractivity (Step 3) transfers the conclusion across the correlated ν-rows at a polynomial cost. Uniform marginals are used where the free and the fixed entries of a column must have the same law: in Step 1, so that a restriction can look like a dictator only at a coordinate i with $\beta_1(i) \geq \frac{1}{2}$, and in Step 3, for the monotonicity of D_j and D_l. They are also used, but only for the constants, in the bound $\nu_{jl} \geq 9\alpha\,\nu_j\nu_l$, which gives $\theta = 9\alpha$ in Step 1 and $\kappa = 9\alpha$ in Step 3; for general marginals one only has $\nu_{jl} \geq \alpha \geq \alpha\,\nu_j\nu_l$. Step 4 holds for every marginal μ_j, with $\frac{\sigma}{3}$ replaced by $\frac{2\sigma}{9}$: if the entries of the j-th column in the rows of $\mathbf{L}$ are replaced by fresh uniform values, then each entry and its replacement have the joint law $\sigma\,u\otimes u + (1-\sigma)\,\mathrm{diag}(\nu_j)$ (u uniform), a symmetric coupling of μ_j with itself whose Markov operator T satisfies $\langle\phi, T\phi\rangle_{\mu_j} = \sigma(\mathbb{E}_u\phi)^2 + (1-\sigma)\mathbb{E}_{\nu_j}\phi^2 \geq 0$ and $\langle\phi,(I-T)\phi\rangle_{\mu_j} = \sigma\,\mathrm{Var}_u\phi \geq \frac{4\sigma}{9}\mathrm{Var}_{\mu_j}\phi$ (as $\mathrm{Var}_u\phi \geq \frac{1}{9}(\max\phi - \min\phi)^2$ and $\mathrm{Var}_{\mu_j}\phi \leq \frac{1}{4}(\max\phi - \min\phi)^2$); applying the resulting spectral gap of $T^{\otimes n}$ to the indicators of the values of f_j and using $\mathrm{dc} \leq 1 - \sum_v p_v^2 \leq 2\,\mathrm{dc}$ gives $\mathbb{E}_{\mathbf{L}} D_j(\mathbf{L}) \geq \frac{2\sigma}{9}\mathrm{dc}(f_j)$. Step 5 holds for every full-support μ. So for a general full-support μ it would suffice to show that, when $\beta_j(k) < \frac{1}{2}$ for all j and k, some f_j is within $\mathrm{poly}(\delta)$ of a constant; Steps 1–4 give this when two marginals are uniform. With fewer uniform marginals, the pointer argument of Step 1 and the monotonicity in Step 3 both fail for the decomposition with π; indeed D_j need not be monotone when μ_j is not uniform. The remedy is to decompose μ with a distribution π' that shares two marginals with μ and is already known to be stable (Lemma 8.64 below). (c) The script `neq3_general_mu_check.py` (directory `numerics`) checks, for random μ (general and with uniform marginals), the decomposition, the bounds $K(a,a') \geq \theta P(a')$ used with Lemma 8.59 ($\theta = 2\alpha$, $3\alpha^2$ and 9α), the inequality of Lemma 8.59 by exact enumeration, the monotonicity of D_j and the bound of Step 4, and Lemma 8.35 with $\kappa = 9\alpha$ and $\kappa = 2\alpha$; it checks both parts of Theorem 8.61 for $n = 2$ against all 13824 generalized polymorphisms (a weak test: there the bounds exceed the distances by many orders of magnitude), the inequalities of Step 5 exactly for $n \leq 2$, the coupling operator T and the bound $\mathbb{E}_{\mathbf{L}} D_j(\mathbf{L}) \geq \frac{2\sigma}{9}\mathrm{dc}(f_j)$ of (b) for non-uniform μ_j (exactly for $n \leq 3$), an instance where D_j is not monotone, and Proposition 8.62(a) for $k \leq 3$ (the error) and $k \leq 2$ (the distance, against all generalized polymorphisms). The script `neq_product_check.py` checks the inequality behind Step 2, in the form $\Pr[f_j \neq u_j]\,\mathrm{dc}(f_s) \leq \mathrm{err}(\mathbf{f}) + \Pr[f_j \neq u_j]\big(\sum_{l\neq j,s}\Pr[f_l \neq u_l] + \Pr[f_s = -\sum_{l\neq s} u_l]\big)$

for all constants u_l ($l \neq s$) and all $j \neq s$, exactly: under π for all triples with $n = 1$ (where it is attained with equality) and for random triples with $n = 2$, and under π_4 (the form used in the proof of Theorem 8.80) for random tuples with $n = 1, 2$. The script `neq3_shift_check.py` checks Proposition 8.62(b): the properties of μ_α^{sh} (exactly); $N_0(\mathbf{y}) - N_1(\mathbf{z}) \leq \mathbf{D}$ for all row tuples with $n \leq 4$; the bounds on the error, exactly, for $n \leq 36$, three values of α and every $\ell \leq s$; the distance bound for $n = 3, 6, 9$ and every $\ell \leq s$ against all generalized polymorphisms, i.e. against the closed-form families (A) and (B), which are complete by Theorem 8.55(a) (with equality in all six cases); the binomial bounds for three values of α and five values of n for each, up to 18324; and the constants.

Every full-support distribution. The decomposition in the proof of Theorem 8.61 can be made with any distribution π' in place of π, provided that π' is already known to be stable and shares two marginals with μ: then the free and the fixed entries of those two columns have the same law, which is what Steps 1 and 3 use. For a distribution ψ on $P_{/=}$ with full support write err_ψ for the error when the rows are independent samples of ψ, and call ψ *(C, c)-stable* if for every n and every triple $\mathbf{f}$ of functions on $\mathbb{Z}_3^n$ there is a generalized polymorphism $\mathbf{g}$ with $\max_t \Pr_{\psi_t^{\otimes n}}[f_t \neq g_t] \leq C\,\mathrm{err}_\psi(\mathbf{f})^c$, where ψ_t is the t-th marginal of ψ (for $n = 0$ the functions are constants, and the condition holds whenever $C \geq 1$). Thus Theorem 8.55(b) says that π is $(4 \cdot 10^7, \frac{1}{2})$-stable, and Theorem 8.61(b) says that ψ is $(3 \cdot 10^7 \alpha^{-3/2}, \frac{rr'}{4})$-stable if two of its marginals are uniform. As $\mathrm{err}_\psi \leq 1$, a (C, c)-stable distribution is (C_1, c_1)-stable whenever $C_1 \geq C$ and $c_1 \leq c$.

Lemma 8.64 (Transfer)**.** *Let μ and π' be distributions on $P_{/=}$ with full support such that, for some $j \neq l$, the j-th marginals of μ and π' coincide and so do their l-th marginals. Let $0 < a \leq \min \mu$ and $0 < a' \leq \min \pi'$, and suppose that π' is (C', c')-stable, where $C' \geq 1$ and $0 < c' \leq 1$. Let $r = r_a$ and $r' = r_{2a}$ (Lemma 8.35; note $2a \leq \frac{1}{9}$), $\theta = \min(2a', \frac{a}{2})$, $\delta_1 = (\theta/C')^{1/c'}$ and $K = C' + \frac{15}{a\delta_1}$. Then μ is (C, c)-stable with*

$$c = \frac{c' r r'}{4}, \qquad C = \max\Big(5 + \frac{16}{a} K^{r/2},\ \frac{16}{a^4}\Big).$$

Proof. Permuting the coordinates preserves $P_{/=}$, so we may assume $\{j, l\} = \{1, 2\}$. Let $\alpha = \min \mu \geq a$, and let $\mathbf{f}$ be a triple with $\mathrm{err}_\mu(\mathbf{f}) = \delta$; note $c \leq 1$. If $\delta > a^4/16$, every generalized polymorphism is within $1 < 16\delta/a^4 \leq 16a^{-4}\delta^c$. If $\delta \leq a^4/16$ and $\beta_t(k) \geq \frac{1}{2}$ for some t and k, Lemma 8.60(a) gives $\mathbf{g}$ with $\mathrm{dist}(\mathbf{f}, \mathbf{g}) \leq 3\delta/\alpha^4 \leq 16a^{-4}\delta^c$. So we may assume that $\beta_t(k) < \frac{1}{2}$ for all t and k.

The decomposition. Let $\sigma = \frac{a}{2}$ and $\nu = (\mu - \sigma\pi')/(1 - \sigma)$. As $\sigma\pi' \leq \frac{a}{2} \leq \frac{\mu}{2}$, ν is a distribution on $P_{/=}$ with $\nu \geq \frac{\mu}{2} \geq \frac{a}{2}$. For $t \in \{1, 2\}$ its t-th marginal is $\nu_t = (\mu_t - \sigma\pi'_t)/(1 - \sigma) = \mu_t$. As every pair has two completions, the law ν_{12} of the first two entries of a ν-row satisfies $\nu_{12}(u, v) \geq a \geq a\,\nu_1(u)\nu_2(v)$. Choose $\mathbf{L}$, ρ and $\mathbf{f}^{L,\rho}$ as in the proof of Theorem 8.61, with π' in place of π: each i is in $\mathbf{L}$ independently with probability σ, the rows in $\mathbf{L}$ are independent samples of π' and those outside are independent samples of ν, so that the rows are independent samples of μ. Given L and ρ, $\mathbf{f}^{L,\rho}$ is a triple of functions on $\mathbb{Z}_3^L$ with error $\delta(L, \rho) := \mathrm{err}_{\pi'}(\mathbf{f}^{L,\rho})$; let $\delta_L = \mathbb{E}_\rho \delta(L, \rho)$, so $\mathbb{E}_{\mathbf{L}} \delta_{\mathbf{L}} = \delta$. For $t \in \{1, 2\}$ the entries of the t-th column have law μ_t both inside and outside L (as $\pi'_t = \nu_t = \mu_t$), so $\rho_t \sim \mu_t^{\otimes L^c}$; below, dc of a function on $\mathbb{Z}_3^L$ obtained from f_t refers to $\mu_t^{\otimes L} = \pi_t'^{\otimes L}$.

Step 1: restricted dictators are rare. For $t \in \{1, 2\}$ let $P_t(\rho_t) = i$ if $f_t(\cdot, \rho_t)$ is θ-close to a function $\kappa w_i + d$ with $i \in L$, $\kappa \in \{1, 2\}$ and $d \in \mathbb{Z}_3$, and $P_t(\rho_t) = \perp$ otherwise. Each value has probability at least $6a'$ under π'_t, as it is the t-th entry of six elements of $P_{/=}$. Hence two distinct functions of this form differ with probability at least $12a' > 2\theta$ (for the same i they differ at at least two values of w_i; for $i \neq i'$ each value of w_i agrees with at most one value of $w_{i'}$, so they agree with probability at most $\max_v \pi'_t(v) \leq 1 - 12a'$), so i is unique; and $\kappa w_i + d$ differs with

probability at least $6a' > 2\theta$ from every function with at most two values. If $P_1(\rho_1) \neq P_2(\rho_2)$, then no generalized polymorphism is θ-close to $\mathbf{f}^{L,\rho}$ under the marginals of π': one of type (B) with coordinate i would give $P_1(\rho_1) = P_2(\rho_2) = i$, and in one of type (A) the first two functions take at most two values each, which would give $P_1(\rho_1) = P_2(\rho_2) = \perp$. As π' is (C', c')-stable, $C'\delta(L,\rho)^{c'} > \theta$, i.e. $\delta(L,\rho) > \delta_1$. By Markov's inequality, $\Pr_\rho[P_1(\rho_1) \neq P_2(\rho_2)] \leq \delta_L/\delta_1$, and by Lemma 8.59 for the first two entries of a ν-row, with a in place of the parameter θ of that lemma (as $\nu_{12} \geq a\,\nu_1\nu_2$), and the argument of fact (ii) in the proof of Lemma 8.60,

$$\mathrm{dc}(P_1(\rho_1)) \leq \eta_L := \frac{2\delta_L}{a\delta_1}.$$

Suppose $\eta_L \leq \frac{1}{7}$ and the most likely value of $P_1(\rho_1)$ is some $i \in L$. Whenever $P_1(\rho_1) = i$, the function $f_1(\cdot, \rho_1)$ differs from some $\kappa w_i + d$ only on a set S with $\mu_1^{\otimes L}(S) \leq \theta$. For $w'' \in \mathbb{Z}_3^{L\setminus\{i\}}$, if $u \mapsto f_1(u, w'', \rho_1)$ is not a bijection, then the line $\{(u, w'') : u \in \mathbb{Z}_3\}$ meets S, at a point whose conditional probability given the line is at least $\min \mu_1 \geq 6a$; so for $\mathbf{w}'' \sim \mu_1^{\otimes(L\setminus\{i\})}$ this map is a bijection except with probability at most $\theta/(6a) \leq \frac{1}{12}$. Since $(\mathbf{w}'', \rho_1) \sim \mu_1^{\otimes([n]\setminus\{i\})}$, this gives $\beta_1(i) \geq \frac{11}{12}(1 - \eta_L) > \frac{1}{2}$, contrary to our assumption. Hence $\Pr_\rho[P_1(\rho_1) \neq \perp] \leq \eta_L$ if $\eta_L \leq \frac{1}{7}$, and $\Pr_\rho[P_1(\rho_1) \neq \perp] \leq 7\eta_L$ in all cases.

Step 2: two nearly constant restricted functions. If $P_1(\rho_1) = \perp$ and $\delta(L,\rho) \leq \delta_1$, the stability of π' gives a generalized polymorphism within $C'\delta(L,\rho)^{c'} \leq \theta$ of $\mathbf{f}^{L,\rho}$. It is not of type (B), as its coordinate would be $P_1(\rho_1)$; so it is of type (A), and two of the three functions of $\mathbf{f}^{L,\rho}$, one of them $f_1(\cdot,\rho_1)$ or $f_2(\cdot,\rho_2)$, are $C'\delta(L,\rho)^{c'}$-close to constants. Hence $m(L,\rho) := \min\big(\mathrm{dc}(f_1(\cdot,\rho_1)), \mathrm{dc}(f_2(\cdot,\rho_2))\big) \leq C'\delta(L,\rho)^{c'}$ in this case, and $m \leq 1$ always. By Step 1, Markov's inequality, Jensen's inequality and $\delta_L \leq \delta_L^{c'}$,

$$\mathbb{E}_\rho\, m(L,\rho) \leq C'\delta_L^{c'} + 7\eta_L + \frac{\delta_L}{\delta_1} \leq K\delta_L^{c'}.$$

Step 3: reverse hypercontractivity and Harris's inequality. Let $D_t(L) = \mathbb{E}_{\rho_t}\mathrm{dc}(f_t(\cdot,\rho_t))$ for $t \in \{1,2\}$. Lemma 8.35 for the independent pairs $(\rho_{1,i}, \rho_{2,i})$, $i \notin L$, with law ν_{12} and $\kappa = a$, gives $D_1(L)D_2(L) \leq K^r\delta_L^{c'r}$, exactly as in Step 3 of the proof of Theorem 8.61, now with $\Phi = \mathrm{dc}(f_1(\cdot,\rho_1))$ and $\Psi = \mathrm{dc}(f_2(\cdot,\rho_2))$, which take values in $[0,1]$, so that $\Phi\Psi \leq \min(\Phi,\Psi) = m(L,\rho)$. There the monotonicity of D_j in L used only that the free and the fixed entries of the j-th column have the same law; here both have law μ_t, so D_1 and D_2 are nondecreasing in L. By Harris's inequality and Jensen's inequality, $\mathbb{E}_{\mathbf{L}}D_1(\mathbf{L}) \cdot \mathbb{E}_{\mathbf{L}}D_2(\mathbf{L}) \leq K^r\delta^{c'r}$.

Step 4: noise. Let $t \in \{1,2\}$, let $\mathbf{U}$ be the t-th column, and let $\mathbf{U}''$ be obtained from $\mathbf{U}$ by replacing its entries in the rows of $\mathbf{L}$ by fresh independent samples of μ_t. The pairs $(\mathbf{U}_i, \mathbf{U}_i'')$ are independent with law $\sigma\,\mu_t \otimes \mu_t + (1-\sigma)\,\mathrm{diag}(\mu_t)$, whose Markov operator $(1-\sigma)I + \sigma\mathbb{E}_{\mu_t}$ has the eigenvalue 1 on constants and $1-\sigma$ on their orthogonal complement. As in the proof of Lemma 8.59, $\Pr[f_t(\mathbf{U}) = f_t(\mathbf{U}'')] \leq 1 - \sigma + \sigma\max_v \Pr[f_t = v] = 1 - \sigma\,\mathrm{dc}(f_t)$. Given $\mathbf{L} = L$ and ρ_t, the probability that $f_t(\mathbf{U}) \neq f_t(\mathbf{U}'')$ is $1 - \sum_v p_v^2 \leq 2\,\mathrm{dc}(f_t(\cdot,\rho_t))$, where $p_v = \Pr[f_t(\cdot,\rho_t) = v]$. Hence $\mathbb{E}_{\mathbf{L}}D_t(\mathbf{L}) \geq \frac{\sigma}{2}\mathrm{dc}(f_t) = \frac{a}{4}\mathrm{dc}(f_t)$.

Conclusion. By Steps 3 and 4, some $t \in \{1,2\}$ has $\mathrm{dc}(f_t) \leq \varepsilon := E_0\delta^{c'r/2}$, where $E_0 = \frac{4}{a}K^{r/2}$. Step 5 of the proof of Theorem 8.61 holds with α replaced by a (Lemma 8.35 applies with $\kappa = 2a$, as $\mu_{lm} \geq 2\alpha \geq 2a\,\mu_l\mu_m$), so a second index has $\mathrm{dc} \leq 2(\delta+\varepsilon)^{r'/2}$, and, as in the conclusion of that proof, some $\mathbf{g}$ of type (A) has $\mathrm{dist}(\mathbf{f},\mathbf{g}) \leq \delta + 2\varepsilon''$, where $\varepsilon'' = \max\big(\varepsilon, 2(\delta+\varepsilon)^{r'/2}\big)$. Since $c \leq \frac{c'r}{2} \leq 1$, we have $\delta \leq \delta^c$, $\varepsilon \leq E_0\delta^c$ and $2(\delta+\varepsilon)^{r'/2} \leq 2(1+E_0)^{r'/2}\delta^{c'rr'/4} \leq 2(1+E_0)\delta^c$. So $\mathrm{dist}(\mathbf{f},\mathbf{g}) \leq (5+4E_0)\delta^c \leq C\delta^c$. □

To apply Lemma 8.64 we connect μ to a distribution with two uniform marginals by distributions that share two marginals at each step.

Lemma 8.65 (A chain of marginals)**.** *Let μ be a distribution on $P_{\not=}$ with full support and $\alpha = \min \mu$. There are distributions π_1, π_2, π_3 on $P_{\not=}$ with $\min \pi_1 \ge \frac{\alpha}{2}$, $\min \pi_2 \ge \frac{\alpha}{30}$ and $\min \pi_3 \ge \frac{\alpha}{2}$ such that π_1 has the second and third marginals of μ, π_2 has the first and third marginals of π_1, π_3 has the second and third marginals of π_2, and the first two marginals of π_3 are uniform.*

Proof. The completions of a pair (u, v) of entries in two columns to an element of $P_{\not=}$ are the two values different from $-u - v$. Let ψ be a distribution on $\mathbb{Z}_3^2$, regarded as the law of the entries in two columns l, m, and let j be the third column. For $e \in \mathbb{Z}_3$ let $T_e = \psi(\{(u, v) : u + v = -e\})$ and $s_e = 1 - T_e$, the mass of the pairs that e completes; so $\sum_e T_e = 1$ and $\sum_e s_e = 2$.

Claim. If d is a distribution on $\mathbb{Z}_3$, $0 < \varepsilon < 1$ and $\frac{\varepsilon}{2} s_e \le d(e) \le (1 - \frac{\varepsilon}{2}) s_e$ for all e, then there is a distribution ψ' on $P_{\not=}$ whose law in the columns l, m is ψ, whose j-th marginal is d, and such that $\psi'(p) \ge \frac{\varepsilon}{2} \psi(p_l, p_m)$ for all $p \in P_{\not=}$.

Indeed, let $\tilde{d} = (d - \frac{\varepsilon}{2} s)/(1 - \varepsilon)$; then $\sum_e \tilde{d}(e) = 1$ and $0 \le \tilde{d}(e) \le s_e = 1 - T_e$. Put $\tau = \max\big(0, \tilde{d}(1) - T_2, T_0 - \tilde{d}(2)\big)$ and

$$x_{01} = \tau,\ \ x_{02} = T_0 - \tau,\ \ x_{21} = \tilde{d}(1) - \tau,\ \ x_{20} = T_2 - \tilde{d}(1) + \tau,\ \ x_{10} = 1 - \tilde{d}(2) - T_2 - \tau,\ \ x_{12} = \tilde{d}(2) - T_0 + \tau.$$

Then $\sum_{a \ne e} x_{ea} = T_e$ for every e and $\sum_{e \ne a} x_{ea} = \tilde{d}(a)$ for every a, and all x_{ea} are nonnegative: each of the nine inequalities between one of the three lower bounds and one of the three upper bounds for τ implicit here reduces to $\tilde{d} \ge 0$, $T \ge 0$, $\sum_a \tilde{d}(a) = \sum_e T_e = 1$ or $\tilde{d}(a) \le 1 - T_a$. (This is the supply–demand theorem for three suppliers, supplier e being excluded from consumer e.) For $p \in P_{\not=}$ with $p_j = a$ and $(p_l, p_m) = (u, v)$, so that $e := -u - v \ne a$, let $\psi'(p) = \psi(u, v)\big(\frac{\varepsilon}{2} + (1 - \varepsilon) x_{ea}/T_e\big)$, and $\psi'(p) = 0$ if $\psi(u, v) = 0$. The two completions of (u, v) get total weight $\psi(u, v)\big(\varepsilon + (1 - \varepsilon) T_e/T_e\big) = \psi(u, v)$, and the j-th marginal of ψ' is $\frac{\varepsilon}{2} s_a + (1 - \varepsilon) \sum_{e \ne a} x_{ea} = d(a)$. This proves the claim.

Each value has probability at least 6α in each marginal of μ, and every pair law of μ is at least 2α. Let m_3 be the third marginal of μ.

π_1. Apply the claim to $\psi = \mu_{23}$ (columns $2, 3$, and $j = 1$), with $\varepsilon = \frac{1}{2}$ and $d = m_1'$, where $m_1'(a) = \min(\frac{2}{3}, s_a)/Z$ and $Z = \sum_b \min(\frac{2}{3}, s_b)$. As Z is a concave function of (s_0, s_1, s_2) on the polytope $\{\sum_e s_e = 2,\ 0 \le s_e \le 1\}$, whose vertices are the permutations of $(1, 1, 0)$, we have $Z \ge \frac{4}{3}$; also $Z \le 2$. Hence $\frac{s_a}{4} \le \frac{s_a}{3} \le \frac{1}{2} \min(\frac{2}{3}, s_a) \le m_1'(a) \le \frac{3}{4} \min(\frac{2}{3}, s_a)$, so the claim applies, $\pi_1 \ge \frac{1}{4} \mu_{23} \ge \frac{\alpha}{2}$, and $m_1'(a) \le \frac{1}{2}$ for all a.

π_2. Let $\pi_{1,3}$ be the law of the first and third entries of a π_1-row, and $\gamma = \frac{5}{6} m_1' \otimes m_3 + \frac{1}{6} \pi_{1,3}$, a distribution on $\mathbb{Z}_3^2$ with marginals m_1' and m_3 and with $\gamma \ge \frac{1}{6} \pi_{1,3} \ge \frac{1}{6} \cdot 2 \cdot \frac{\alpha}{2} = \frac{\alpha}{6}$. For each b, $\gamma(\{(a, c) : a + c = -b\}) \le \frac{5}{6} \max_a m_1'(a) + \frac{1}{6} \le \frac{7}{12}$. So with $\psi = \gamma$ (columns $1, 3$, and $j = 2$) we have $s_b \ge \frac{5}{12}$, and the claim applies with $\varepsilon = \frac{2}{5}$ and d uniform, as $\frac{1}{5} s_b \le \frac{1}{3} \le \frac{4}{5} \cdot \frac{5}{12} \le \frac{4}{5} s_b$. So $\pi_2 \ge \frac{1}{5} \gamma \ge \frac{\alpha}{30}$, its first and third marginals are m_1' and m_3, and its second marginal is uniform.

π_3. Let Q_0 be the uniform distribution on the six elements $(1, 1, 0)$, $(2, 2, 0)$, $(0, 1, 0)$, $(0, 2, 0)$, $(1, 0, 0)$, $(2, 0, 0)$ of $P_{\not=}$; its first two marginals are uniform. For $c \in \mathbb{Z}_3$ let Q_c be the image of Q_0 under $(a, b, 0) \mapsto (a - c, b, c)$, which preserves the sum of the entries; so Q_c is a distribution on $P_{\not=}$ with uniform first two marginals and third entry c. As $m_3 \ge 6\alpha$ and $9\alpha \le \frac{1}{2}$, $\tilde{m}_3 = (m_3 - 3\alpha)/(1 - 9\alpha)$ is a distribution on $\mathbb{Z}_3$, and

$$\pi_3 = (1 - 9\alpha) \sum_c \tilde{m}_3(c)\, Q_c + 9\alpha\, \pi$$

has uniform first two marginals, third marginal $(m_3 - 3\alpha) + 3\alpha = m_3$, and $\pi_3 \ge 9\alpha/18 = \frac{\alpha}{2}$. □

Theorem 8.66 (Every full-support distribution)**.** *Let μ be a distribution on $P_{\not=}$ with full support and $\alpha = \min \mu$. Then μ is (C_α, c_α)-stable, where*

$$c_\alpha = \frac{\alpha^8}{13107200}, \qquad C_\alpha = \exp\Big(1.2 \cdot 10^{11}\, \alpha^{-9} \ln \frac{1}{\alpha}\Big).$$

That is, for every n and every triple $\mathbf{f}$ of functions on $\mathbb{Z}_3^n$ with $\mathrm{err}_\mu(\mathbf{f}) = \delta$ there is a generalized polymorphism $\mathbf{g}$ with $\mathrm{dist}(\mathbf{f},\mathbf{g}) \le C_\alpha \delta^{c_\alpha}$, distances being measured under $\mu_j^{\otimes n}$.

Proof. Let π_1, π_2, π_3 be as in Lemma 8.65, $x = 1/\alpha \ge 18$ and $\ell = \ln x$. As $\min \pi_3 \ge \frac{\alpha}{2}$ and the first two marginals of π_3 are uniform, Theorem 8.61(b) shows that π_3 is (C_3, c_3)-stable with $C_3 = 3 \cdot 10^7 (2x)^{3/2}$ and $c_3 = \frac{9}{64}\alpha^2 \le \frac{9}{16}(\min \pi_3)^2$. We apply Lemma 8.64 three times, using $r_a \ge \frac{a}{2}$ and $r_{2a} \ge a$, so that the exponent is multiplied by at least $\frac{a^2}{8}$ at each step:

(i) to $\pi' = \pi_3$ and the distribution π_2 (common second and third marginals), with $a = \frac{\alpha}{30}$ and $a' = \frac{\alpha}{2}$: π_2 is (C_2, c_2)-stable with $c_2 = \frac{c_3 \alpha^2}{7200}$;

(ii) to $\pi' = \pi_2$ and π_1 (common first and third marginals), with $a = \frac{\alpha}{2}$ and $a' = \frac{\alpha}{30}$: π_1 is (C_1, c_1)-stable with $c_1 = \frac{c_2 \alpha^2}{32}$;

(iii) to $\pi' = \pi_1$ and μ (common second and third marginals), with $a = \alpha$ and $a' = \frac{\alpha}{2}$: μ is (C, c)-stable with $c = \frac{c_1 \alpha^2}{8} = \frac{9}{64 \cdot 7200 \cdot 32 \cdot 8}\alpha^8 = c_\alpha$.

For the constants: in Lemma 8.64, $\delta_1 \le \theta/C' \le 1/C'$, so $K \le (1 + \frac{15}{a})/\delta_1$; and $r_a \le 2a$, as $1 - \sqrt{1-a} \le a$. Hence

$$\ln C \le \max\Big(\ln \frac{17}{a} + a \ln K,\ \ln \frac{16}{a^4} \Big), \qquad \ln \frac{1}{\delta_1} = \frac{\ln C' + \ln(1/\theta)}{c'}.$$

The lower-order terms are absorbed as follows; two of these inequalities have little room at $x = 18$, and all of them are checked on a grid of $x \in [18, 10^{15}]$ by `numerics/neq_all_constants_check.py`. In each case the difference of the two sides is increasing in $x \ge 18$ (the side that is linear or polynomial in x has the larger derivative), so it suffices to check $x = 18$.

(i) $a = \frac{1}{30x}$, $\theta = \frac{1}{60x}$, $c' = c_3 = \frac{9}{64x^2}$ and $\ln C_3 = \ln(3 \cdot 10^7 \cdot 2^{3/2}) + 1.5\ell \le 18.26 + 1.5\ell$, so $\ln \frac{1}{\delta_1} \le \frac{64}{9}(18.26 + \ln 60 + 2.5\ell)x^2 \le (159 + 17.78\ell)x^2$ and $a \ln K \le \frac{\ln(1+450x)}{30x} + (5.3 + 0.5927\ell)x$. Hence

$$\ln \frac{17}{a} + a \ln K \le 6.24 + \ell + \frac{\ln(1+450x)}{30x} + (5.3 + 0.5927\ell)x \le (5.65 + 0.65\ell)x,$$

as $6.24 + \ell + \frac{\ln(1+450x)}{30x} \le (0.35 + 0.0573\ell)x$ (at $x = 18$: $9.148 \le 9.281$; the left side has derivative at most $\frac{1}{18}$, the right side at least 0.57). Also $\ln \frac{16}{a^4} = \ln 16 + 4\ln(30x) \le 16.38 + 4\ell \le (5.65 + 0.65\ell)x$. So $\ln C_2 \le (5.65 + 0.65\ell)x$.

(ii) $a = \frac{1}{2x}$, $\theta = \frac{1}{15x}$, $c' = c_2 = \frac{1}{51200x^4}$, so

$$\ln \frac{1}{\delta_1} \le 51200\big((5.65 + 0.65\ell)x + \ln 15 + \ell\big)x^4 \le (2.97 \cdot 10^5 + 3.62 \cdot 10^4 \ell)x^5,$$

as $51200 \cdot 5.65 = 289280$, $51200 \cdot 0.65 = 33280$ and $51200(\ln 15 + \ell) \le (7720 + 2920\ell)x$ (at $x = 18$: $286639 \le 290878$; derivatives at most 2845 and at least 19079). Then $\ln \frac{17}{a} + a \ln K \le \ln(34x) + \frac{\ln(1+30x)}{2x} + (1.485 \cdot 10^5 + 1.81 \cdot 10^4 \ell)x^4 \le (1.49 \cdot 10^5 + 1.81 \cdot 10^4 \ell)x^4$, as $3.53 + \ell + \frac{\ln(1+30x)}{2x} \le 500x^4$, and $\ln \frac{16}{a^4} = \ln 16 + 4\ln(2x)$ is smaller. So $\ln C_1 \le (1.49 \cdot 10^5 + 1.81 \cdot 10^4 \ell)x^4$.

(iii) $a = \frac{1}{x}$, $\theta = \frac{1}{2x}$, $c' = c_1 = \frac{1}{1638400x^6}$, so

$$\ln \frac{1}{\delta_1} \le 1638400\big((1.49 \cdot 10^5 + 1.81 \cdot 10^4 \ell)x^4 + \ln 2 + \ell\big)x^6 \le (2.442 \cdot 10^{11} + 2.966 \cdot 10^{10} \ell)x^{10},$$

as $1638400 \cdot 1.49 \cdot 10^5 < 2.4413 \cdot 10^{11}$, $1638400 \cdot 1.81 \cdot 10^4 < 2.9656 \cdot 10^{10}$ and $1638400(\ln 2 + \ell) \le (7 \cdot 10^7 + 4 \cdot 10^6 \ell)x^4$. Then $\ln \frac{17}{a} + a \ln K \le \ln(17x) + \frac{\ln(1+15x)}{x} + (2.442 \cdot 10^{11} + 2.966 \cdot 10^{10}\ell)x^9 \le (2.45 \cdot 10^{11} + 2.97 \cdot 10^{10}\ell)x^9$, and $\ln \frac{16}{a^4} = \ln 16 + 4\ell$ is smaller. Finally $(2.45 \cdot 10^{11} + 2.97 \cdot 10^{10}\ell)x^9 \le 1.2 \cdot 10^{11} \ell x^9$ iff $2.45 \cdot 10^{11} \le 9.03 \cdot 10^{10}\ell$, which holds as $\ell \ge \ln 18 > 2.89$.

□

Remark 8.67. (a) Theorem 8.66 answers the first question of [2, Section 9] for $P_{/=}$: under every full-support distribution, approximate generalized polymorphisms are close to exact ones, with polynomial dependence. By Proposition 8.62 the exponent must depend on μ, and by Proposition 8.82 no exponent larger than 9α holds for every full-support μ with $\min \mu = \alpha$. The exponent c_α and the constant C_α are far from optimal; for distributions with uniform marginals, Theorem 8.61(a) gives the exponent $r_{9\alpha}/2 \ge 4\alpha$ with a moderate constant, which is of the optimal order α for this class, and Proposition 8.81 gives every exponent below $r_{18\alpha}/2 \ge 8\alpha$, within a factor $\frac{9}{8}$ of the upper bound 9α for this class (and every exponent below $32\alpha^2$ when only two marginals are uniform). (b) Some intermediate distributions are needed, as Lemma 8.64 requires two common marginals at each step. For instance, a distribution on $P_{/=}$ with uniform first marginal and the second and third marginals of μ need not exist: if μ is close to the point mass at $(0, 0, 1)$, its second and third marginals are close to the point masses at 0 and 1, while a row $(a, 0, 1) \in P_{/=}$ has $a \in \{0, 1\}$. Lemma 8.65 first replaces the first marginal by $m_1' \le \frac{1}{2}$, which leaves room to make the second marginal uniform, and only then makes the first marginal uniform. (c) Theorem 8.61(b) is close to Lemma 8.64 with $\pi' = \pi$, which is stable by Theorem 8.55(b). The differences are the constants, the parameters and Step 2: there $\sigma = 9\alpha$ (as $\pi \le \frac{1}{18}$) and reverse hypercontractivity is applied with $\kappa = 9\alpha$, which uses the uniform marginals, while Lemma 8.64 has $\sigma = \frac{a}{2}$ and $\kappa = a$; and Step 2 there bounds the product of the two distances to constants, using the pairwise independence of the columns under π, while Lemma 8.64 bounds only their minimum. So its exponent $r_a r_{2a}/8$ (for $a = \alpha$) is smaller than $r_{9\alpha} r_{2\alpha}/4$ by a factor of about 18. (d) The script `neq3_transfer_check.py` (directory `numerics`) checks Lemma 8.65 in exact rational arithmetic for 1989 distributions, among them distributions near each point mass, near two- and three-point distributions, and with one or two nearly degenerate marginals: the transportation solution, all marginals, and the bounds on $\min \pi_1$, $\min \pi_2$, $\min \pi_3$ (the smallest observed ratios to α are $\frac{1}{2}$, $\frac{1}{6}$ and 1). For the three steps of the proof of Theorem 8.66 it checks the decomposition of Lemma 8.64, the common marginals, the hypotheses of Lemma 8.35 with $\kappa = a$ and $\kappa = 2a$, and Lemma 8.35 itself by enumeration. It checks the monotonicity of D_t and the bound of Step 4 when the free and the fixed entries have the same non-uniform law, and the facts about $\kappa w_i + d$ and the bijection bound used in Step 1, exactly for $n \le 3$; and it checks the arithmetic of Theorem 8.66 for $\alpha \in [10^{-12}, \frac{1}{18}]$.

The second question of [2, Section 9] asks, for Boolean predicates, for the additional guarantee that $g_j = g_l$ whenever $f_j = f_l$ and $\mu_j = \mu_l$. For $P_{/=}$ this guarantee follows from Theorem 8.66 by the routine argument that [2, Section 9] give for the predicate $\{(a, b, a \wedge b)\}$ (replace a function close to a constant by that constant), and for polymorphisms with $\mu_1 = \mu_2 = \mu_3$ the dependence becomes linear; under π the sharp constant is $\frac{2}{3}$.

Corollary 8.68 (Equal functions and polymorphisms)**.** *Let μ be a distribution on $P_{/=}$ with full support, $\alpha = \min \mu$, C_α and c_α as in Theorem 8.66, and* $\mathrm{err}_\mu(\mathbf{f}) = \delta$.

(a) If $\mu_j = \mu_l$ and $f_j = f_l$ for some $j \ne l$, then there is a generalized polymorphism $\mathbf{g}$ with $g_j = g_l$ and $\mathrm{dist}(\mathbf{f}, \mathbf{g}) \le 2C_\alpha \delta^{c_\alpha}$. If moreover $n \ge 1$, $\mu_1 = \mu_2 = \mu_3$ and $f_1 = f_2 = f_3$, then $\mathbf{g}$ can be taken with $g_1 = g_2 = g_3$. So $g_j = g_l$ can be achieved simultaneously for all pairs with $\mu_j = \mu_l$ and $f_j = f_l$ (these pairs are the pairs of an equivalence class).

(b) *A function $h : \mathbb{Z}_3^n \to \mathbb{Z}_3$ is a polymorphism of $P_{\not=}$, i.e. (h,h,h) is a generalized polymorphism, iff $h = \kappa w_i + d$ for some $i \in [n]$, $\kappa \in \{1,2\}$ and $d \in \mathbb{Z}_3$. In substance this is due to Greenwell and Lovász [14] (see also [1]): the triple (x,y,y) is a row of $P_{\not=}$ in every coordinate iff $x_i \neq y_i$ for all i, and then $h(x) + 2h(y) \neq 0$ iff $h(x) \neq h(y)$; so h is a proper 3-colouring of the n-th direct power of K_3, hence $h = \rho(w_i)$ for a permutation ρ of $\mathbb{Z}_3$, and every permutation of $\mathbb{Z}_3$ is affine.*

(c) *Let $n \geq 1$. If $\mu_1 = \mu_2 = \mu_3 =: \nu$ and $f_1 = f_2 = f_3 =: h$, then $\Pr_{\nu^{\otimes n}}[h \neq \kappa w_i + d] \leq K_\mu \delta$ for some i, κ and d, where $K_\mu = (C_\alpha/\alpha)^{1/c_\alpha}$. For $\mu = \pi$ and $\delta \leq 10^{-15}$ the bound holds with $K_\mu\delta$ replaced by $\frac{2\delta}{3-12\delta}$.*

(d) *For $\mu = \pi$ the constant $\frac{2}{3}$ in (c) is sharp: if $n \geq 2$, $E \subseteq \mathbb{Z}_3^n$ is determined by $w_2, \dots, w_n$, $s = \Pr[E] \leq \frac{1}{3}$ and $h = w_1 + \mathbf{1}_E$, then $\operatorname{err}(h,h,h) = \frac{3}{2}s(1-s)$, while $\Pr[h \neq \kappa w_i + d] \geq s$ for all i, κ and d, with equality for w_1.*

Proof. Throughout, $\Pr_\nu$ stands for $\Pr_{\nu^{\otimes n}}$. We use three facts. First, by Theorem 8.55(a) every generalized polymorphism is of type (A) or (B). Second, if $f_j = f_l =: h$, $\mu_j = \mu_l =: \nu$ and $\Pr_\nu[h \neq g_j], \Pr_\nu[h \neq g_l] \leq \eta < \frac{1}{2}$, then $\Pr_\nu[g_j \neq g_l] \leq 2\eta < 1$; in particular, g_j, g_l are not two distinct constants, and if $g_j = \kappa w_k + d_j$ and $g_l = \kappa w_k + d_l$, then $d_j = d_l$, since two such functions with $d_j \neq d_l$ differ everywhere. Third, if $f_1 = f_2 = f_3 =: h$, $\mu_1 = \mu_2 = \mu_3 =: \nu$ and $\operatorname{dist}(\mathbf{f}, \mathbf{g}) \leq \eta < \frac{1}{2}$ for a generalized polymorphism $\mathbf{g}$, then $\mathbf{g}$ is of type (B) and $g_1 = g_2 = g_3$. Indeed, if $\mathbf{g}$ were of type (A) with constants u, v at positions a, b and a function g_c avoiding $-u - v$, then $u = v$ by the second fact, so $g_c \neq -2u = u$ everywhere, and $1 - \eta \leq \Pr_\nu[h = u] \leq \Pr_\nu[h \neq g_c] \leq \eta$, which is false. So $\mathbf{g}$ is of type (B), and $g_1 = g_2 = g_3$ by the second fact.

(a) Let $\mathbf{g}$ be given by Theorem 8.66 and $\eta = \operatorname{dist}(\mathbf{f}, \mathbf{g}) \leq C_\alpha \delta^{c_\alpha}$; write $h = f_j = f_l$, $\nu = \mu_j = \mu_l$, and let m be the third index. If $\eta \geq \frac{1}{2}$, the type-(A) triple with $g'_j = g'_l \equiv 0$ and $g'_m \equiv 1$ satisfies $\operatorname{dist}(\mathbf{f}, \mathbf{g}') \leq 1 \leq 2\eta$. Let $\eta < \frac{1}{2}$. If $\mathbf{g}$ is of type (B), then $g_j = g_l$ by the second fact. Let $\mathbf{g}$ be of type (A), with constants u, v at positions a, b and a function avoiding $-u - v$ at position c. If $\{j, l\} = \{a, b\}$, then $u = v$ by the second fact, so $g_j = g_l$. Otherwise one of j, l is c and the other, say a, carries the constant u; by the second fact g_c takes the value u, so $u \neq -u - v$, i.e. $v \neq u$ (as $-2u = u$). Then $\mathbf{g}'$ with $g'_a = g'_c \equiv u$ and $g'_b = g_b \equiv v$ is of type (A) (its entry at position b, the constant v, avoids $-2u = u$), $g'_j = g'_l$, and $\operatorname{dist}(\mathbf{f}, \mathbf{g}') \leq \eta$, because $\Pr_\nu[h \neq u] \leq \eta$ is the distance at both positions a and c. Now let $n \geq 1$, $\mu_1 = \mu_2 = \mu_3$ and $f_1 = f_2 = f_3$. If $\eta \geq \frac{1}{2}$, the triple (w_1, w_1, w_1), of type (B), is within distance $1 \leq 2\eta$; if $\eta < \frac{1}{2}$, then $g_1 = g_2 = g_3$ by the third fact. Finally, the pairs $\{j, l\}$ with $\mu_j = \mu_l$ and $f_j = f_l$ are the pairs of an equivalence relation on $\{1, 2, 3\}$, so there are none, exactly one, or all three of them.

(b) Every $\kappa w_i + d$ is a polymorphism, since $(\kappa x_i + d) + (\kappa y_i + d) + (\kappa z_i + d) = \kappa(x_i + y_i + z_i) \neq 0$. Conversely, let (h, h, h) be a generalized polymorphism. It is not of type (A): then h would be a constant u and the third entry $h \equiv u$ would avoid $-2u = u$. So it is of type (B), $h = \kappa w_i + d_j$ for $j = 1, 2, 3$, and $d_1 = d_2 = d_3$.

(c) Let $\delta_0 = (\alpha/C_\alpha)^{1/c_\alpha} = K_\mu^{-1}$. As $C_\alpha \geq 1$ and $c_\alpha \leq \frac{1}{8}$, $\delta_0 \leq \alpha^8 \leq \alpha^4/16$. If $\delta > \delta_0$, then $K_\mu\delta > 1$ and there is nothing to prove. Let $\delta \leq \delta_0$, so $C_\alpha \delta^{c_\alpha} \leq \alpha$, and let $\mathbf{g}$ be given by Theorem 8.66: $\operatorname{dist}(\mathbf{f}, \mathbf{g}) \leq \alpha < \frac{1}{2}$. By the third fact $\mathbf{g}$ is of type (B), with some coordinate k. By Lemma 8.60(b) (note $\delta \leq \alpha^4/16$) there is $\mathbf{g}'$ of type (B), $g'_l = \kappa w_k + d_l$, with $\Pr_\nu[h \neq g'_l] \leq 3\delta/\alpha^4 \leq \frac{3}{16}$ for $l = 1, 2, 3$. By the second fact $d_1 = d_2 = d_3 =: d$, and $\Pr_\nu[h \neq \kappa w_k + d] \leq 3\delta/\alpha^4 \leq K_\mu\delta$, as $K_\mu = \delta_0^{-1} \geq 16/\alpha^4$.

For $\mu = \pi$ and $\delta \leq 10^{-15}$, Theorem 8.55(b) gives $\mathbf{g}$ with $\operatorname{dist}(\mathbf{f}, \mathbf{g}) \leq 3\sqrt{\delta} < \frac{1}{10}$, which is of type (B) by the third fact (with $\eta = 3\sqrt{\delta}$), and Theorem 8.55(d) (note $\delta \leq 8 \cdot 10^{-5}$) gives $\mathbf{g}'$ of

type (B) with $\sum_j \Pr[h \neq g'_j] \leq 3\delta$. By the second fact the g'_j are equal, say to $p = \kappa w_i + d$, and $s := \Pr[h \neq p] \leq \delta$. The first inequality of the display in the proof of Theorem 8.55(d) uses only that $\mathbf{g}'$ is of type (B) (not the bound $\delta'_j \leq 882\delta$); for the triple (p, p, p), which is of type (B) as $3d = 0$, it reads $\delta \geq \frac{3}{2}s - 6s^2$. As $s \leq \delta$, this gives $\delta \geq s(\frac{3}{2} - 6\delta)$, i.e. $s \leq \frac{2\delta}{3-12\delta}$.

(d) Let $S = x_1 + y_1 + z_1$ and $N = \mathbf{1}_E(x) + \mathbf{1}_E(y) + \mathbf{1}_E(z)$. Then S is uniform on $\{1, 2\}$ (each nonzero value is the sum of 9 elements of $P_{/=}$) and independent of N, which depends only on rows $2, \ldots, n$, and $h(x) + h(y) + h(z) = S + N$. Let $p_t = \Pr[N = t]$. Any two columns are independent and uniform (Lemma 8.56(a)), so $\mathbb{E}N = 3s$ and $\mathbb{E}\binom{N}{2} = 3s^2$, i.e. $p_1 + 2p_2 + 3p_3 = 3s$ and $p_2 + 3p_3 = 3s^2$; hence $p_1 + p_2 = 3s - 3s^2$. As $S \neq 0$, a failure needs $N \in \{1, 2\}$, and given such N it has conditional probability $\frac{1}{2}$. So $\mathrm{err} = \frac{1}{2}(p_1 + p_2) = \frac{3}{2}s(1 - s)$. Finally $\Pr[h \neq w_1] = s$, and every other $\kappa w_i + d$ agrees with w_1 with probability at most $\frac{1}{3}$, so it differs from h with probability at least $\frac{2}{3} - s \geq s$. □

So for polymorphisms under π the supremum of the possible distances at error at most δ (over all n) is $(\frac{2}{3} + O(\delta))\delta$ as $\delta \to 0$. By (c) it is at most $\frac{2\delta}{3-12\delta}$. By (d) it is at least s_δ, the smaller root of $\frac{3}{2}s(1 - s) = \delta$, because s in (d) can be any multiple of 3^{1-n} in $(0, \frac{1}{3}]$ and these are dense; and $s_\delta = \frac{2\delta}{3(1-s_\delta)} \geq \frac{2}{3}\delta$. The script `neq3_poly_check.py` (directory `numerics`) checks (b) for $n \leq 2$ by exhaustive search (6 and 12 polymorphisms); the error formula and the distance claim of (d) for $2 \leq n \leq 4$ by exact enumeration, for $E = \{w_2 = \cdots = w_n = 0\}$ and for random sets E determined by $w_2, \ldots, w_n$; and the inequality $\delta \geq \frac{3}{2}s - 6s^2$ of (c), for every $p = \kappa w_i + d$, all h with $n = 1$ and random h with $n = 2, 3$.

More summands. For $m \geq 3$ let $P^{(m)}_{/} == \{a \in \mathbb{Z}_3^m : a_1 + \cdots + a_m \neq 0\}$, so $P^{(3)}_{/} == P_{/=}$, and let π_m be the uniform distribution on $P^{(m)}_{/} =$. As for $m = 3$, no coordinate is flexible (changing a_j to $-\sum_{l \neq j} a_l$ leaves the predicate), so [2, Theorem 1.10] does not apply. Let $\mathbf{c}_1, \ldots, \mathbf{c}_m$ be the columns of n independent samples of π_m, $\mathrm{err}(\mathbf{f}) = \Pr[\sum_j f_j(\mathbf{c}_j) = 0]$ for $\mathbf{f} = (f_1, \ldots, f_m)$, and $\mathrm{dist}(\mathbf{f}, \mathbf{g}) = \max_j \Pr[f_j \neq g_j]$ under the uniform measure; generalized polymorphisms are the $\mathbf{g}$ with $\mathrm{err}(\mathbf{g}) = 0$. A tuple $\mathbf{g}$ is of *type* (A_m) if for some b the functions g_j, $j \neq b$, are constants u_j and g_b never takes the value $-\sum_{j \neq b} u_j$, and of *type* (B_m) if $g_j = \kappa w_i + d_j$ for all j, with $\kappa \in \{1, 2\}$, a common $i \in [n]$ and $\sum_j d_j = 0$. Both are generalized polymorphisms; for (B_m) the sum is κ times the sum of row i.

Lemma 8.69. *Let $m \geq 4$.*

(a) Any $m - 1$ of the columns are independent and uniform on $\mathbb{Z}_3^n$. Given $\mathbf{c}_j$ (for any j), the row sums $S_i = \sum_l \mathbf{c}_{l,i}$, $i \in [n]$, are independent and uniform on $\{1, 2\}$.

(b) Let $a, b, c \in [m]$ be distinct, $R = [m] \setminus \{a, b, c\}$ and $\mathbf{s} = \sum_{j \in R} \mathbf{c}_j$. Then the columns $\mathbf{c}_j$, $j \in R$, are independent of $(\mathbf{c}_a + \mathbf{s}, \mathbf{c}_b, \mathbf{c}_c)$, which has the law of the three columns of n independent samples of π.

Proof. The rows are independent, so it suffices to consider one row $(y_1, \ldots, y_m) \sim \pi_m$. (a) Any $m - 1$ entries can be completed to an element of $P^{(m)}_{/} =$ in exactly two ways, so they are uniform. Given $y_j = t$, the other entries are uniform on the tuples whose sum is not $-t$; the sum of a uniform tuple in $\mathbb{Z}_3^{m-1}$ is uniform, so given $y_j = t$ the row sum is uniform on $\{1, 2\}$. (b) The map that adds $\sum_{j \in R} y_j$ to the entry y_a and leaves the other entries unchanged is a bijection of $\mathbb{Z}_3^m$, and it maps $P^{(m)}_{/} =$ onto $\{y' : y'_a + y'_b + y'_c \neq 0\}$, which is $P_{/=}$ in the coordinates a, b, c times $\mathbb{Z}_3^R$. So it maps π_m to the product of π and the uniform distribution on $\mathbb{Z}_3^R$. □

Theorem 8.70 (m summands)**.** *Let $m \ge 4$ and $\delta_m = \min\{5\cdot 10^{-16}, (6\sqrt{2}(m-1))^{-4}\}$.*

(a) Every generalized polymorphism of $P_{/}^{(m)}$ = is of type (A_m) or (B_m).

(b) For every n and every $\mathbf{f}$ with $\mathrm{err}(\mathbf{f}) = \delta$ there is a generalized polymorphism $\mathbf{g}$ of type (A_m) or (B_m) with $\mathrm{dist}(\mathbf{f},\mathbf{g}) \le \delta_m^{-1/2}\sqrt{\delta}$, and with $\mathrm{dist}(\mathbf{f},\mathbf{g}) \le \sqrt{\delta}+(2+12(m-1)^2)\delta \le (1+2m\,\delta^{1/4})\sqrt{\delta}$ if $\delta \le \delta_m$.

(c) The constant 1 in (b) is optimal: if $E \subseteq \mathbb{Z}_3^n$ has $\Pr[E] = \eta \le \frac{1}{3}$, $f_1 = \cdots = f_{m-2} \equiv 0$, $f_{m-1} = \mathbf{1}_E$ and $f_m = 1 + \mathbf{1}_E$, then $\mathrm{err}(\mathbf{f}) = \eta^2$ and $\mathrm{dist}(\mathbf{f},\mathbf{g}) \ge \eta$ for every generalized polymorphism $\mathbf{g}$.

(d) Near (B_m) the dependence is linear: if $\delta \le \delta_m$ and $\mathrm{dist}(\mathbf{f},\mathbf{g}) \le \frac{1}{10}$ for some $\mathbf{g}$ of type (B_m), then $\sum_j \Pr[f_j \ne g'_j] \le 3\delta$ for some $\mathbf{g}'$ of type (B_m).

Proof. Let $\delta = \mathrm{err}(\mathbf{f}) \le \delta_m$, $\theta = 3\sqrt{\delta}$ and $\delta_j = \mathrm{dc}(f_j)$. As $\delta \le 5\cdot 10^{-16}$ and $6\sqrt{2}(m-1)\delta^{1/4} \le 1$, we have $\theta < \frac{1}{10}$ and $(m-1)\theta = 3(m-1)\delta^{1/4}\cdot\delta^{1/4} \le \frac{1}{2}\delta^{1/4} \le \frac{1}{12}$. Recall that a function $\kappa w_i + d$ differs from every constant with probability $\frac{2}{3}$ and from every other function $\kappa' w_{i'} + d'$ with probability at least $\frac{2}{3}$; so a function that is θ-close to some $\kappa w_i + d$ is θ-close to no other such function, and its distance to the constants exceeds θ.

Step 1: triples. Let $a,b,c \in [m]$ be distinct and R, $\mathbf{s}$ as in Lemma 8.69(b). For $w = (w_j)_{j\in R}$ put $s(w) = \sum_{j\in R} w_j$, $K(w) = \sum_{j\in R} f_j(w_j)$ and $F^w(x) = f_a(x - s(w)) + K(w)$. Then $\sum_j f_j(\mathbf{c}_j) = F^{\mathbf{w}}(\mathbf{c}_a + \mathbf{s}) + f_b(\mathbf{c}_b) + f_c(\mathbf{c}_c)$ for $\mathbf{w} = (\mathbf{c}_j)_{j\in R}$, so by Lemma 8.69(b), $\delta = \mathbb{E}\,\delta(\mathbf{w})$, where $\delta(w)$ is the error of the triple (F^w, f_b, f_c) for $P_{/=}$ under π. Choose w with $\delta(w) \le \delta$. As $\delta \le 10^{-15}$, Theorem 8.55(b) gives a generalized polymorphism (g_a, g_b, g_c) of $P_{/=}$ of type (A) or (B) with $\Pr[F^w \ne g_a], \Pr[f_b \ne g_b], \Pr[f_c \ne g_c] \le 3\sqrt{\delta(w)} \le \theta$. Note that $\mathrm{dc}(F^w) = \delta_a$, since $x - s(w)$ is uniform when x is. Hence, for every pair $b \ne c$: either f_b and f_c are θ-close to $\kappa w_i + d_b$ and $\kappa w_i + d_c$ for some κ, i, d_b, d_c (type (B)), or $\delta_b \le \theta$ or $\delta_c \le \theta$ (type (A): two of g_a, g_b, g_c are constants, and one of them is g_b or g_c).

Step 2: if $\delta_b, \delta_c > \theta$ for some $b \ne c$, then $\mathbf{f}$ is close to type (B_m). By Step 1, f_b and f_c are θ-close to $\kappa w_i + d_b$ and $\kappa w_i + d_c$. Let $j \notin \{b,c\}$. If $\delta_j \le \theta$, apply Step 1 to $(a,b,c) = (j,b,c)$: then F^w is θ-close to a constant and f_b, f_c are not, so the triple (g_j, g_b, g_c) is neither of type (A) (two constants) nor of type (B) (F^w would be θ-close to some $\kappa' w_{i'} + d'$), a contradiction. So $\delta_j > \theta$, and Step 1 for the pair $\{b,j\}$ shows that f_b and f_j are θ-close to $\kappa' w_{i'} + d'$ and $\kappa' w_{i'} + d_j$; by uniqueness $(\kappa', i', d') = (\kappa, i, d_b)$. Hence every f_j is θ-close to $g_j := \kappa w_i + d_j$. Here $\sum_j g_j(\mathbf{c}_j) = \kappa S_i + \sum_j d_j$, where the row sum $S_i = \sum_j \mathbf{c}_{j,i}$ is uniform on $\{1,2\}$ (Lemma 8.69(a)); if $\sum_j d_j \ne 0$, then $\mathrm{err}(\mathbf{g}) = \frac{1}{2}$ and $\delta \ge \frac{1}{2} - m\theta$, which is false. So $\mathbf{g}$ is of type (B_m). Let $\delta'_j = \Pr[f_j \ne g_j] \le \theta$, $E'_j = \{f_j \ne g_j\}$ and $e_j = f_j - g_j$. On the event that $\mathbf{c}_j \in E'_j$ and $\mathbf{c}_l \notin E'_l$ for all $l \ne j$, the sum $\sum_l f_l(\mathbf{c}_l)$ equals $e_j(\mathbf{c}_j) + \kappa S_i$. Given $\mathbf{c}_j$, S_i is uniform on $\{1,2\}$, so $\Pr[\mathbf{c}_j \in E'_j,\ e_j(\mathbf{c}_j) + \kappa S_i = 0] = \frac{1}{2}\delta'_j$, and removing the events $\{\mathbf{c}_l \in E'_l\}$, $l \ne j$, costs at most $\delta'_j \sum_{l\ne j} \delta'_l$ by pairwise independence. These m events are disjoint, so

$$\delta \ge \frac{1}{2}\sum_j \delta'_j - \sum_j \delta'_j \sum_{l\ne j} \delta'_l \ge \Big(\frac{1}{2} - (m-1)\theta\Big)\sum_j \delta'_j \ge \frac{1}{3}\sum_j \delta'_j. \tag{15}$$

Step 3: otherwise $\mathbf{f}$ is close to type (A_m). Suppose that $\delta_j > \theta$ for at most one j; relabel so that $\delta_j \le \theta$ for $j < m$. Let τ be a number with $2(m-1)\theta \le \tau \le \frac{1}{4m}$. Let u_j be the most likely value of f_j ($j < m$), $U = \sum_{j<m} u_j$, $\Delta = \sum_{j<m} \delta_j \le (m-1)\theta \le \frac{\tau}{2}$, $p_j(t) = \Pr[f_j = u_j + t]$, $r_v = \Pr[f_m = v]$ and $r^* = r_{-U}$. The event that $f_j(\mathbf{c}_j) = u_j$ for all $j < m$ and $f_m(\mathbf{c}_m) = -U$, and for each $j < m$ the event that $f_j(\mathbf{c}_j) = u_j + t \ne u_j$, $f_l(\mathbf{c}_l) = u_l$ for $l < m$, $l \ne j$, and

$f_m(\mathbf{c}_m) = -U - t$, are disjoint failures. By pairwise independence their probabilities are at least $r^*(1-\Delta)$ and $\sum_{t\neq 0} p_j(t) r_{-U-t} - \delta_j \Delta$, so

$$\delta \geq (1-\Delta) r^* + \sum_{j<m} \sum_{t\neq 0} p_j(t) r_{-U-t} - \Delta^2. \tag{16}$$

Case (α): $r_{-U+1}, r_{-U+2} \geq \tau$. Then the double sum is at least $\tau\Delta$, and as $\Delta \leq \frac{\tau}{2} \leq \frac{1}{16}$, (16) gives $\delta \geq \frac{15}{16} r^* + \frac{\tau}{2}\Delta$. Let $g_j = u_j$ for $j < m$, and $g_m = f_m$ except that $g_m = -U+1$ where $f_m = -U$; $\mathbf{g}$ is of type (A_m) and $\mathrm{dist}(\mathbf{f}, \mathbf{g}) \leq \max(\Delta, r^*) \leq 2\delta/\tau$. Case ($\beta$): otherwise. If $r^* \geq \tau$, then $\delta \geq (1-\Delta) r^* \geq \frac{15}{16}\tau$. If $r^* < \tau$, then f_m takes some value $u_m \neq -U$ with probability more than $1 - 2\tau$, so $\delta_m < 2\tau$, and all δ_j are at most 2τ (as $\theta \leq \tau$). Let $\sigma = \sum_{j\leq m} u_j = U + u_m \neq 0$, and for each $j \in [m]$ let $D_j = \{f_j = u_j - \sigma\}$, $A_j = \{f_j = u_j + \sigma\}$, $M_j = \{f_j = u_j\}$, $p_j^* = \Pr[D_j]$ and $p_j^\circ = \Pr[A_j]$, so $\delta_j = p_j^* + p_j^\circ$. The events $\{D_j(\mathbf{c}_j),\ M_l(\mathbf{c}_l)\ \forall l \neq j\}$ ($j \in [m]$) and $\{A_j(\mathbf{c}_j),\ A_l(\mathbf{c}_l),\ M_k(\mathbf{c}_k)\ \forall k \notin \{j,l\}\}$ ($j < l$) are disjoint failures (the sums are $\sigma - \sigma$ and 3σ). As any three columns are independent (Lemma 8.69(a), $m \geq 4$), their probabilities are at least $p_j^*(1 - 2(m-1)\tau)$ and $p_j^\circ p_l^\circ (1 - 2(m-2)\tau)$. Hence, if q is the second largest of $p_1^\circ, \dots, p_m^\circ$,

$$\delta \geq \big(1 - 2(m-1)\tau\big)\Big(\sum_j p_j^* + \sum_{j<l} p_j^\circ p_l^\circ\Big) \geq \big(1 - 2(m-1)\tau\big)\Big(\sum_j p_j^* + q^2\Big).$$

Let j_0 be an index at which $p_{j_0}^\circ$ is largest, $g_j = u_j$ for $j \neq j_0$, and $g_{j_0} = f_{j_0}$ except that $g_{j_0} = u_{j_0}$ on D_{j_0}. Then g_{j_0} never takes the value $u_{j_0} - \sigma = -\sum_{j\neq j_0} u_j$, so $\mathbf{g}$ is of type (A_m), and $\mathrm{dist}(\mathbf{f}, \mathbf{g}) \leq \sum_j p_j^* + q \leq \delta' + \sqrt{\delta'}$ with $\delta' = \delta/(1 - 2(m-1)\tau)$.

Conclusion. If $\delta = 0$, then $\theta = 0$; take $\tau = \frac{1}{4m}$. Step 2 gives $\mathbf{f} = \mathbf{g}$ by (15); in Step 3, $\delta \geq \frac{15}{16}\tau$ is impossible and the other bounds are 0. So $\mathbf{f}$ is of type (A_m) or (B_m), which proves (a). If $0 < \delta \leq \delta_m$, take $\tau = 2(m-1)\theta = 6(m-1)\sqrt{\delta}$; this is allowed, as $6(m-1)\delta^{1/4} \leq \frac{1}{\sqrt{2}}$, so $\tau \leq \delta^{1/4}/\sqrt{2} \leq \frac{1}{12(m-1)} \leq \frac{1}{4m}$. Step 2 gives $\mathrm{dist}(\mathbf{f}, \mathbf{g}) \leq 3\delta$. In Step 3, $\frac{15}{16}\tau \geq 16\sqrt{\delta} > \delta$, so the alternative $\delta \geq \frac{15}{16}\tau$ is impossible; Case (α) gives $2\delta/\tau = \sqrt{\delta}/(3(m-1))$; and in Case ($\beta$), with $x = 2(m-1)\tau = 12(m-1)^2\sqrt{\delta} \leq \frac{1}{6}$ (as $\sqrt{\delta} \leq \delta_m^{1/2} \leq (72(m-1)^2)^{-1}$), we have $(1-x)^{-1} \leq 2$ and $(1-x)^{-1/2} \leq 1 + x$, so the bound is at most $2\delta + (1+x)\sqrt{\delta} = \sqrt{\delta} + (2 + 12(m-1)^2)\delta$. As 3δ and $\sqrt{\delta}/(3(m-1))$ are smaller, this proves the second bound in (b). Moreover $(2 + 12(m-1)^2)\delta^{1/4} \leq \frac{2+12(m-1)^2}{6\sqrt{2}(m-1)} \leq \sqrt{2}(m-1) + \frac{1}{4} \leq 2m$, so this bound is at most $(1 + 2m\delta^{1/4})\sqrt{\delta}$; and $(2 + 12(m-1)^2)\sqrt{\delta} \leq \frac{2+12(m-1)^2}{72(m-1)^2} < 1$, so it is at most $2\sqrt{\delta} \leq \delta_m^{-1/2}\sqrt{\delta}$. If $\delta > \delta_m$, the tuple $(0, \dots, 0, 1)$, which is of type (A_m), is within distance $1 \leq \delta_m^{-1/2}\sqrt{\delta}$. This proves (b).

(c) A failure needs $\mathbf{1}_E(\mathbf{c}_{m-1}) + 1 + \mathbf{1}_E(\mathbf{c}_m) \equiv 0 \pmod 3$, i.e. $\mathbf{c}_{m-1}, \mathbf{c}_m \in E$, so $\mathrm{err}(\mathbf{f}) = \eta^2$ by Lemma 8.69(a). If $\mathbf{g}$ is of type (B_m), then $\Pr[f_1 \neq g_1] = \frac{2}{3} \geq \eta$. If $\mathbf{g}$ is of type (A_m), then g_{m-1} or g_m is a constant, and $\mathbf{1}_E$ and $1 + \mathbf{1}_E$ differ from every constant with probability at least $\min(\eta, 1-\eta) = \eta$.

(d) Each f_j is $\frac{1}{10}$-close to some $\kappa w_i + d_j$, so $\delta_j \geq \frac{2}{3} - \frac{1}{10} > \theta$ for all j. So Step 2 applies and gives $\mathbf{g}'$ of type (B_m) with $\sum_j \Pr[f_j \neq g'_j] \leq 3\delta$ by (15). □

Remark 8.71. (a) With Theorem 8.55, the predicates $P_{/}^{(m)} =$, $m \geq 3$, all have the optimal exponent $\frac{1}{2}$ and the sharp constant 1 under the uniform distribution. For $m \geq 4$ the bound in (b) is $\sqrt{\delta} + O(m^2\delta)$, while Theorem 8.55(b) gives $(1 + O(\delta^{1/8}))\sqrt{\delta}$ for $m = 3$. The difference comes from Step 1, which obtains the distance $\theta = O(\sqrt{\delta})$ directly from Theorem 8.55(b) and so allows the threshold $\tau = 2(m-1)\theta$ in Step 3; the independence of any three columns only removes the

three-wise term X of Case (β) of the proof of Theorem 8.55. In (d) the constant 3 can be replaced by $\frac{2}{1-6\delta}$, as for $m=3$: for the $\mathbf{g}'$ of (d), $S=\sum_j \Pr[f_j \neq g'_j] \leq 3\delta$, and the first inequality in (15), which holds for every tuple of type (B_m) in place of $\mathbf{g}$, gives $\delta \geq \frac{S}{2} - S^2 \geq (\frac{1}{2}-3\delta)S$ (the second inequality, with $\theta = 3\sqrt{\delta}$, gives $\frac{2}{1-6(m-1)\sqrt{\delta}}$). The constant 2 is optimal: changing one function of a tuple $\mathbf{g}$ of type (B_m) on a set E of measure $\eta \leq \frac{1}{3}$ gives error exactly $\frac{\eta}{2}$ (off E the sum $\sum_j f_j(\mathbf{c}_j)$ is $\kappa S_i \neq 0$, and given the changed column, the row sum S_i is uniform on $\{1,2\}$), while $\sum_j \Pr[f_j \neq g'_j] \geq \eta$ for every $\mathbf{g}'$ of type (B_m), as for $m=3$; for $\eta \leq 2\delta_m$ this tuple satisfies the hypotheses of (d). (b) Part (a) can also be reduced to the case $n=2$ by [10, Theorem 1.3], as for $m=3$; we do not need this, since (a) follows from (b). (c) Other full-support distributions μ on $P_{/}^{(m)}=$, $m \geq 4$, are not covered by the reduction above: given the columns in R, the rows of $(\mathbf{c}_a+\mathbf{s}, \mathbf{c}_b, \mathbf{c}_c)$ are independent but not identically distributed, so Theorem 8.66 does not apply to them directly. They are treated by a decomposition of μ in Theorem 8.78 below, which uses the following extension of Lemma 8.60(a), with err_μ, dist, $\beta_j(k)$ and dc defined as for $m=3$ (before Lemma 8.59). First, Lemma 8.60 holds for $P_{/}^{(m)}=$, with type (B_m) in place of (B), $l \in [m]$, the hypothesis $\delta \leq \alpha^4/(4(m+1))$ and the bound $2\delta/\alpha^4$; so near (B_m) the dependence is linear for every such μ. In its proof, (i) is unchanged; in (ii) every (a,b) has $2\cdot 3^{m-3} \geq 2$ completions; in (iii), with $\mathbf{v}$ the other $m-1$ entries, at least 3^{m-2} values v are completed by both a and a' (if $a \neq a'$, those with $\sum v = -a''$), so Lemma 8.59 applies with $\theta = 3^{m-2}\alpha^2$; in (a), Theorem 8.70(a) for $n=1$ replaces Lemma 8.57 (no function of a tuple of type (A_m) is a bijection, every function of a tuple of type (B_m) is), $\xi \leq m\delta/\alpha^2 + \delta/\alpha \leq (m+1)\delta/\alpha^2 \leq \alpha^2/4$, the copy is taken given the other $m-1$ columns, $\mathrm{dc}(d_l(\mathbf{c}'_l)) \leq 2\xi/(3^{m-2}\alpha^2) \leq \frac{1}{2\cdot 3^{m-2}}$, so that $\Pr[\sum_l d_l(\mathbf{c}'_l)=0,\ d_l(\mathbf{c}'_l)=d_l^0\ \forall l] \geq 1-\xi-\frac{m}{2\cdot 3^{m-2}} > 0$, and the final bound is $\frac{\delta}{\alpha^2} + \frac{(m+1)\delta}{\alpha^3} + \frac{2(m+1)\delta}{3^{m-2}\alpha^4} \leq \frac{2\delta}{\alpha^4}$, as $\alpha \leq \frac{1}{2\cdot 3^{m-1}}$; in (b), $\mu_j(a) \geq 2\cdot 3^{m-2}\alpha \geq 6\alpha$. Second, as for $m=3$ (Proposition 8.62), the exponent of a polynomial bound for these distributions, such as that of Theorem 8.78, must depend on μ. For $0<\alpha\leq\frac{1}{2\cdot 3^{m-1}}$ let $\mu_m^{(\alpha)}$ give mass α to each of the $2\cdot 3^{m-2}$ elements of $P_{/}^{(m)}=$ with $a_{m-1}=a_m$ and mass $\frac{1-2\cdot 3^{m-2}\alpha}{4\cdot 3^{m-2}} \geq \alpha$ to each of the other $4\cdot 3^{m-2}$ elements; for $m=3$ this is $\mu^{(\alpha)}$ of Proposition 8.62, and $\mu_m^{(\alpha)}=\pi_m$ for $\alpha=\frac{1}{2\cdot 3^{m-1}}$. It is invariant under the translations $a \mapsto a+t$ with $\sum_j t_j = 0$ and $t_{m-1}=t_m$, and for each j and $s \in \mathbb{Z}_3$ one of them shifts the j-th entry by s (for $j \leq m-2$ take $t_j = t_{m-1} = t_m = s$, for $j \geq m-1$ take $t_1 = t_{m-1} = t_m = s$, and the other entries 0); so its marginals are uniform, and $\min \mu_m^{(\alpha)} = \alpha$. For $k \geq 1$ and $E=\{0\} \subseteq \mathbb{Z}_3^k$, the tuple of Theorem 8.70(c) fails exactly when $\mathbf{c}_{m-1}=\mathbf{c}_m=0$, and a row has $y_{m-1}=y_m=0$ with probability $2\cdot 3^{m-3}\alpha$, so its error is $(2\cdot 3^{m-3}\alpha)^k$; the proof of Theorem 8.70(c) uses only (a), which concerns only the support, and the uniformity of the columns, so its distance from every generalized polymorphism is at least 3^{-k}. Hence if every tuple (for every n) is within $C\,\mathrm{err}_{\mu_m^{(\alpha)}}^c$ of a generalized polymorphism, then $c \leq \log 3/\log\frac{1}{2\cdot 3^{m-3}\alpha}$, which tends to 0 as $\alpha \to 0$. (d) The script `neqm_check.py` (directory `numerics`) checks Lemma 8.69 exactly for $m=4,5,6$; part (a) for $n=1$, $m=4$ by exhaustive search (756 generalized polymorphisms: 702 of type (A_4) and 54 of type (B_4)), and for the same n and m the two bounds of (b) for all 27^4 tuples, with exact errors and distances; the example of (c) for $m=4$, $n\leq 3$ and $m=5$, $n\leq 2$; the two bounds of (b) by exact enumeration on random perturbations of generalized polymorphisms for $m=4$, $n=2,3$ and $m=5$, $n=2$ (the largest ratio $\mathrm{dist}/\sqrt{\delta}$ in this sample is 1, but this is specific to the sample: for $n=1$, $m=4$ the largest ratio over all tuples is $\frac{2}{3}/\sqrt{5/18} \approx 1.265$, at $\delta=\frac{5}{18}$); and the bound of (d) on the 19 such perturbations with $m=4$, $n=3$ that satisfy its hypothesis (for $n=2$ the hypothesis forces $\mathbf{f}$ to be of type (B_m)). These tests have δ far above δ_m, so they test the inequalities outside the range in which they are claimed. It also checks the distribution $\mu_m^{(\alpha)}$ of (c) exactly: its marginals and minimum for $m=3,4,5$, and the error and the distance bound of

its example for $(m,k) = (4,1),(4,2),(5,1)$.

The largest predicates without flexible coordinates. If no coordinate of $P \subseteq \mathbb{Z}_3^m$ is flexible, then no axis-parallel line $\{w + te_j : t \in \mathbb{Z}_3\}$ is contained in P (any point of such a line would witness that coordinate j is flexible). The 3^{m-1} lines in direction m partition $\mathbb{Z}_3^m$, so $|P| \le 2 \cdot 3^{m-1}$, with equality for $P_{\not=}^{(m)}$. We show that $P_{\not=}^{(m)}$ is the only predicate attaining this, up to relabelling the alphabet in each coordinate. For bijections $\sigma_1, \dots, \sigma_m$ of $\mathbb{Z}_3$ write $\sigma(a) = (\sigma_1(a_1), \dots, \sigma_m(a_m))$, and for $\mathbf{f} = (f_1, \dots, f_m)$, $f_j \colon \mathbb{Z}_3^n \to \mathbb{Z}_3$, let $(\sigma \cdot \mathbf{f})_j = \sigma_j \circ f_j \circ \sigma_j^{-1}$, where σ_j^{-1} acts on each of the n coordinates; for a distribution μ on $\mathbb{Z}_3^m$ let $\sigma_*\mu$ be the law of $\sigma(\mathbf{y})$ for $\mathbf{y} \sim \mu$. We use the following classical fact; we include the short proof for completeness.

Lemma 8.72 ([20, §4], [27, §1], [19, p. 224])**.** *Let $k \ge 1$ and let $L \colon \mathbb{Z}_3^k \to \mathbb{Z}_3$ be a bijection in each argument when the other arguments are fixed. Then $L(a) = \varepsilon_1 a_1 + \cdots + \varepsilon_k a_k + c$ for some $\varepsilon_1, \dots, \varepsilon_k \in \{1,2\}$ and $c \in \mathbb{Z}_3$.*

Proof. For $k = 1$: the six permutations of $\mathbb{Z}_3$ are the six maps $t \mapsto \varepsilon t + c$, $\varepsilon \in \{1,2\}$. For $k = 2$: by the case $k = 1$, $L(a,b) = \varepsilon_a b + c_a$. If $\varepsilon_a \ne \varepsilon_{a'}$ for some $a \ne a'$, then $L(a,b) - L(a',b) = (\varepsilon_a - \varepsilon_{a'})b + c_a - c_{a'}$ vanishes for exactly one b, as $\varepsilon_a - \varepsilon_{a'} \ne 0$; this contradicts the injectivity of $a \mapsto L(a,b)$. So $\varepsilon_a = \varepsilon$ for all a, hence $a \mapsto c_a = L(a,b) - \varepsilon b$ is a bijection, and $c_a = \varepsilon' a + c$ by the case $k = 1$. For $k \ge 3$, by induction, $L(t, a_2, \dots, a_k) = \sum_{j \ge 2} \varepsilon_j(t) a_j + c(t)$ for each $t \in \mathbb{Z}_3$. Fix $j \ge 2$ and the arguments a_l, $l \notin \{1, j\}$. Then $(t, a_j) \mapsto L$ satisfies the hypothesis with $k = 2$, and its row t has slope $\varepsilon_j(t)$ in a_j; by the case $k = 2$ all rows have the same slope, so $\varepsilon_j(t) = \varepsilon_j$. Now $t \mapsto L(t, a_2, \dots, a_k) = c(t) + \sum_{j \ge 2} \varepsilon_j a_j$ is a bijection, so $c(t) = \varepsilon_1 t + c$ by the case $k = 1$. □

Corollary 8.73 (Maximum predicates without flexible coordinates)**.** *Let $m \ge 3$ and let $P \subseteq \mathbb{Z}_3^m$ have no flexible coordinate and $|P| = 2 \cdot 3^{m-1}$. Then $P = \{a : \varepsilon_1 a_1 + \cdots + \varepsilon_m a_m \ne c\}$ for some $\varepsilon_j \in \{1,2\}$ and $c \in \mathbb{Z}_3$, so $P = \sigma(P_{\not=}^{(m)})$ for some bijections σ_j of $\mathbb{Z}_3$. For every such σ and every distribution μ on $P_{\not=}^{(m)}$, the map $\mathbf{f} \mapsto \sigma \cdot \mathbf{f}$ is a bijection from tuples for $(P_{\not=}^{(m)}, \mu)$ to tuples for $(P, \sigma_*\mu)$ that preserves the error and the distances under the column marginals, and it maps generalized polymorphisms onto generalized polymorphisms. Consequently Theorems 8.55 ($m = 3$) and 8.70 ($m \ge 4$) hold verbatim for P under the uniform distribution on P, and Theorems 8.66 ($m = 3$) and 8.78 ($m \ge 4$, below) for every full-support distribution on P, with the types (A), (B), (A_m), (B_m) and the example tuples of part (c) replaced by their images under σ.*

Proof. Every axis-parallel line contains at most two points of P. The lines in direction j partition $\mathbb{Z}_3^m$ into 3^{m-1} lines, so $|P| = 2 \cdot 3^{m-1}$ forces every line, in every direction, to contain exactly two points of P; that is, $Q = \mathbb{Z}_3^m \setminus P$ meets every axis-parallel line in exactly one point. So for each $a' \in \mathbb{Z}_3^{m-1}$ there is exactly one $L(a')$ with $(a', L(a')) \in Q$. If a', a'' differ only in coordinate $j < m$ and $L(a') = L(a'')$, then $(a', L(a'))$ and $(a'', L(a''))$ are two points of Q on one line in direction j, which is impossible; so L is injective, hence bijective, in each argument. By Lemma 8.72 ([20, 27, 19]) with $k = m - 1$, $L(a') = \sum_{j<m} \varepsilon_j a_j + c_0$, so $a \in P$ iff $\sum_{j<m} \varepsilon_j a_j + 2a_m + c_0 \ne 0$. With $\tau_j(t) = \varepsilon_j t$ for $j < m$ and $\tau_m(t) = 2t + c_0$ we get $P = \tau^{-1}(P_{\not=}^{(m)})$; take $\sigma = \tau^{-1}$.

Let the rows of an $n \times m$ matrix be independent samples of μ, with columns $\mathbf{c}_j$. Applying σ to every row gives independent samples of $\sigma_*\mu$, whose columns are $\mathbf{c}'_j = \sigma_j(\mathbf{c}_j)$ (coordinatewise), and $(\sigma \cdot \mathbf{f})_j(\mathbf{c}'_j) = \sigma_j(f_j(\mathbf{c}_j))$. So the row of values of $\sigma \cdot \mathbf{f}$ is σ applied to the row of values of $\mathbf{f}$, and it lies in P iff the latter lies in $P_{\not=}^{(m)}$. This gives equality of the errors, and, applied to all matrices with rows in $P_{\not=}^{(m)}$, the correspondence of generalized polymorphisms. Also $\Pr[(\sigma \cdot \mathbf{f})_j(\mathbf{c}'_j) \ne (\sigma \cdot \mathbf{g})_j(\mathbf{c}'_j)] = \Pr[f_j(\mathbf{c}_j) \ne g_j(\mathbf{c}_j)]$, and $\sigma^{-1} \cdot (\sigma \cdot \mathbf{f}) = \mathbf{f}$. Finally σ_* is a bijection

from the distributions on $P^{(m)}_{/=}$ onto the distributions on P, with inverse $(\sigma^{-1})_*$; it maps the uniform distribution to the uniform distribution (whose marginals are uniform), and full-support distributions to full-support distributions with the same smallest atom. So for a tuple $\mathbf{f}$ for P the three theorems, applied to $\sigma^{-1}\cdot\mathbf{f}$, give the statements for $\mathbf{f}$, and the image under σ of an example tuple of part (c) has the same error and the same distance to the generalized polymorphisms. □

Remark 8.74. (a) So the predicate that [2, Section 9] single out is, up to relabelling the alphabet in each coordinate, the unique largest predicate in $\mathbb{Z}_3^3$ without a flexible coordinate, and Theorems 8.55, 8.66, 8.70 and 8.78 cover all largest predicates without flexible coordinates over $\mathbb{Z}_3$, for every $m \geq 3$ and every full-support distribution, with the optimal exponent $\frac{1}{2}$ under the uniform distribution. The images of the types do not depend on σ: for $P = \{\sum_j \varepsilon_j a_j \neq c\}$, the image of (A_m) (or (A)) consists of the tuples in which all functions but one, g_b, are constants u_j and g_b never takes the unique value t with $\varepsilon_b t + \sum_{j\neq b}\varepsilon_j u_j = c$, and the image of (B_m) (or (B)) consists of the tuples $g_j = \kappa w_i + d_j$ with a common $\kappa \in \{1,2\}$ and $i \in [n]$ and $\sum_j \varepsilon_j d_j = (1-\kappa)c$. (b) Lemma 8.72 is classical: the k-ary quasigroups of order 3 are exactly the $3\cdot 2^k$ maps $\sum_j \varepsilon_j a_j + c$ (conversely, each such map is a quasigroup, and distinct (ε, c) give distinct maps) [20, §4], [27, §1], [19, p. 224]. A set $Q \subseteq \mathbb{Z}_3^m$ that meets every axis-parallel line exactly once is a Latin hypercube (a ternary MDS code of length m with minimum distance 2), so the first assertion of Corollary 8.73 follows directly from classical facts; what is new is only its connection with the flexibility condition of [2, Theorem 1.10] and the transfer of Theorems 8.55, 8.66, 8.70 and 8.78. (c) The first assertion of Corollary 8.73 and its proof also hold for $m = 2$, with the case $k = 1$ of Lemma 8.72: the largest predicates in $\mathbb{Z}_3^2$ without a flexible coordinate are the sets $\{a_2 \neq \rho(a_1)\}$ for permutations ρ of $\mathbb{Z}_3$, i.e. the relabellings of $\{a_1 \neq a_2\}$. The polymorphisms $(f_1 = f_2)$ of $\{a_1 \neq a_2\}$ are the proper 3-colourings of the n-th direct power of the triangle K_3, and Greenwell and Lovász [14] showed that each of them has the form $x \mapsto \rho(x_i)$. For $m = 2$ approximate versions are known as well: see Alon, Dinur, Friedgut and Sudakov [1] and Ghandehari and Hatami [13] (stability of large independent sets in powers of complete graphs) and Filmus [9]. They are not treated here, and neither are predicates without flexible coordinates that have fewer than $2\cdot 3^{m-1}$ points (for example proper subsets of $P_{/=}$ with full projections). (d) The script `maxnonflex_check.py` (directory `numerics`) enumerates all k-ary quasigroups of order 3 for $k \leq 4$ ($3\cdot 2^k$ of them, all affine with unit coefficients); finds all 18-point predicates in $\mathbb{Z}_3^3$ without a flexible coordinate by a search over all $\binom{27}{9}$ nine-point complements, which does not use the counting argument (12, all of the form $\sum_j \varepsilon_j a_j \neq c$); checks the correspondence for the relabelling $P' = \{a_1 + a_2 + 2a_3 \neq 1\}$ of $P_{/=}$ (198 generalized polymorphisms on each side for $n = 1$; equal errors for 200 random triples with $n = 2$); and checks that for $n = 1$ the generalized polymorphisms of P' are exactly the tuples of the two forms described in (a).

Equal functions and polymorphisms for m summands. We return to $P^{(m)}_{/=}$, $m \geq 4$, under π_m, with the notation of Theorem 8.70. A *polymorphism* is a function $h\colon \mathbb{Z}_3^n \to \mathbb{Z}_3$ such that $(h,\dots,h)$ is a generalized polymorphism.

Corollary 8.75 (Equal functions and polymorphisms, m summands)**.** *Let $m \geq 4$ and $n \geq 1$, and let δ_m be as in Theorem 8.70.*

(a) For every $\mathbf{f}$ with $\mathrm{err}(\mathbf{f}) = \delta$ there is a generalized polymorphism $\mathbf{g}$ of type (A_m) or (B_m) such that $g_j = g_l$ whenever $f_j = f_l$, with $\mathrm{dist}(\mathbf{f},\mathbf{g}) \leq \delta_m^{-1/2}\sqrt{\delta}$, and with $\mathrm{dist}(\mathbf{f},\mathbf{g}) \leq \sqrt{\delta} + (2 + 12(m-1)^2)\delta$ if $\delta \leq \delta_m$.

(b) *The polymorphisms are the functions* $\kappa w_i + d$ *with* $md = 0$ *and the constants* u *with* $mu \neq 0$*: if* $3 \mid m$*, the* $6n$ *functions* $\kappa w_i + d$*; if* $3 \nmid m$*, the* $2n$ *functions* κw_i *and the constants* 1 *and* 2*.*

(c) *Let* $h\colon \mathbb{Z}_3^n \to \mathbb{Z}_3$ *and* $\delta = \operatorname{err}(h,\dots,h) \le \delta_m$*. Some polymorphism* p *satisfies* $\Pr[h \neq p] \le \frac{2\delta}{m(1-6\delta)}$ *if* $3 \mid m$*, and* $\Pr[h \neq p] \le (1+2m\sqrt{\delta})\sqrt{2\delta/(m(m-1))}$ *if* $3 \nmid m$*.*

(d) *Both constants are optimal. Let* $E \subseteq \mathbb{Z}_3^n$ *have* $\Pr[E] = s \le \frac{1}{3}$*, and let* N *be the number of* j *with* $\mathbf{c}_j \in E$*. If* $3 \mid m$ *and* E *is determined by* $w_2,\dots,w_n$*, then* $h = w_1 + \mathbf{1}_E$ *has* $\operatorname{err}(h,\dots,h) = \frac{1}{2}\Pr[N \not\equiv 0 \pmod 3] \le \frac{m}{2}s$*. If* $3 \nmid m$ *and* $u \in \{1,2\}$*, then* $h = u + mu\,\mathbf{1}_E$ *has* $\operatorname{err}(h,\dots,h) = \Pr[N \equiv 2 \pmod 3] \le \binom{m}{2}s^2$*. In both cases* $\Pr[h \neq p] \ge s$ *for every polymorphism* p*, with equality for* $p = w_1$ *and* $p = u$*, respectively.*

Proof. (a) If $\delta > \delta_m$, let $g_1 = \cdots = g_m = w_1$, a tuple of type (B_m) for every m (all $d_j = 0$); its distance from $\mathbf{f}$ is at most $1 \le \delta_m^{-1/2}\sqrt{\delta}$. Let $\delta \le \delta_m$, let $\mathbf{g}$ be given by Theorem 8.70(b), and let $\eta = \operatorname{dist}(\mathbf{f},\mathbf{g}) \le \sqrt{\delta} + (2+12(m-1)^2)\delta$; then $\eta \le 2\sqrt{\delta} < \frac{1}{2}$, as shown in the proof of Theorem 8.70. Call j and l equivalent if $f_j = f_l$; for equivalent j,l we have $\Pr[g_j \neq g_l] \le 2\eta < 1$. If $\mathbf{g}$ is of type (B_m), this forces $g_j = g_l$, since $\kappa w_i + d_j$ and $\kappa w_i + d_l$ differ everywhere if $d_j \neq d_l$. Let $\mathbf{g}$ be of type (A_m), with constants u_j $(j \neq b)$ and g_b avoiding $v = -\sum_{j\neq b} u_j$. Equivalent $j,l \neq b$ carry equal constants, since distinct constants differ everywhere. If b is equivalent to no other index, $\mathbf{g}$ already has the required property. If b is equivalent to some $l \neq b$, then g_b takes the value u_l somewhere, as $\Pr[g_b \neq u_l] < 1$, so $u_l \neq v$. Replacing g_b by the constant u_l gives a tuple of type (A_m) (its entry at position b avoids v) in which equivalent indices carry equal functions, and its distance from $\mathbf{f}$ is still at most η, since $\Pr[f_b \neq u_l] = \Pr[f_l \neq u_l] \le \eta$.

(b) Since $\sum_j(\kappa\mathbf{c}_{j,i} + d) = \kappa S_i + md$ and the row sum S_i takes both values 1 and 2, the function $\kappa w_i + d$ is a polymorphism iff $md = 0$; a constant u is one iff $mu \neq 0$. Conversely, if $(h,\dots,h)$ is a generalized polymorphism, it is of type (A_m) or (B_m) by Theorem 8.70(a). In the first case h is constant, since $m-1 \ge 1$ of the entries are constants; in the second, $h = \kappa w_i + d_j$ for every j, so $d_1 = \cdots = d_m$.

(c) By (a) and its proof, applied to $(h,\dots,h)$, there is a polymorphism p with $\Pr[h \neq p] \le 2\sqrt{\delta} < \frac{1}{10}$. Suppose first that $p = \kappa w_i + d$. Then $(p,\dots,p)$ is of type (B_m), and Theorem 8.70(d) gives $\mathbf{g}'$ of type (B_m) with $\sum_j \Pr[h \neq g'_j] \le 3\delta$. Any two g'_j are 3δ-close, so they are all equal, to p' say, and p' is a polymorphism. Let $s = \Pr[h \neq p'] \le \frac{3\delta}{m}$. The first inequality in (15) holds for every tuple of type (B_m) in place of $\mathbf{g}$ (its derivation uses nothing else), so for $(h,\dots,h)$ and $(p',\dots,p')$ it reads $\delta \ge \frac{m}{2}s - m(m-1)s^2 \ge \frac{m}{2}s(1-6\delta)$, as $2(m-1)s \le 6\delta$. Hence $s \le \frac{2\delta}{m(1-6\delta)}$. This proves (c) if $3 \mid m$, where every polymorphism is of this form, and also if $3 \nmid m$ and p is not constant, since $\frac{2\delta}{m(1-6\delta)} \le \sqrt{2\delta/(m(m-1))}$ (this is equivalent to $2(m-1)\delta \le m(1-6\delta)^2$, which holds as $\delta \le 5\cdot 10^{-16}$). Now let $3 \nmid m$ and $p = u \in \{1,2\}$. Let $\zeta = \Pr[h \neq u] \le 2\sqrt{\delta}$, $q_t = \Pr[h = u + t]$, $t^* = -mu$ and $t^{**} = mu$, so that $\{t^*, t^{**}\} = \{1,2\}$. The events that $h(\mathbf{c}_j) = u + t^*$ for one j and $h(\mathbf{c}_l) = u$ for all $l \neq j$ (the sum is $mu + t^* = 0$), and that $h(\mathbf{c}_j) = h(\mathbf{c}_l) = u + t^{**}$ for one pair $j < l$ and $h(\mathbf{c}_k) = u$ for all other k (the sum is $3mu = 0$), are disjoint failures. Any three columns are independent (Lemma 8.69(a)), so their probabilities are at least $q_{t^*}(1-(m-1)\zeta)$ and $q_{t^{**}}^2(1-(m-2)\zeta)$. With $x = (m-1)\zeta \le 2(m-1)\sqrt{\delta} \le \frac{1}{36(m-1)}$ (as $\sqrt{\delta} \le (72(m-1)^2)^{-1}$) this gives

$$\delta \ge (1-x)\Big(m\,q_{t^*} + \binom{m}{2}q_{t^{**}}^2\Big),$$

so $\zeta = q_{t^*} + q_{t^{**}} \le \frac{\delta}{m(1-x)} + \sqrt{\frac{2\delta}{m(m-1)(1-x)}} \le \frac{2\delta}{m} + (1+x)\sqrt{\frac{2\delta}{m(m-1)}}$. As $\frac{2\delta}{m} = \sqrt{\frac{2(m-1)\delta}{m}}\sqrt{\frac{2\delta}{m(m-1)}} \le \sqrt{2\delta}\sqrt{\frac{2\delta}{m(m-1)}}$, we get $\zeta \le (1+2(m-1)\sqrt{\delta}+\sqrt{2\delta})\sqrt{2\delta/(m(m-1))} \le (1+2m\sqrt{\delta})\sqrt{2\delta/(m(m-1))}$.

(d) Let $3 \mid m$ and $h = w_1 + \mathbf{1}_E$. Then $\sum_j h(\mathbf{c}_j) = S_1 + N$, where S_1 depends only on the first row and is uniform on $\{1,2\}$, and N depends only on the other rows. So a failure has conditional probability $\frac{1}{2}$ given $N \not\equiv 0$ and 0 otherwise, and $\mathrm{err} = \frac{1}{2}\Pr[N \not\equiv 0] \le \frac{1}{2}\mathbb{E}N = \frac{m}{2}s$, as each column is uniform. Also $\Pr[h \ne w_1] = s$, and every other $\kappa w_i + d$ agrees with w_1 with probability at most $\frac{1}{3}$, hence differs from h with probability at least $\frac{2}{3} - s \ge s$. Let $3 \nmid m$, $u \in \{1,2\}$ and $h = u + mu\,\mathbf{1}_E$. Then $\sum_j h(\mathbf{c}_j) = mu(1+N)$, which vanishes iff $N \equiv 2 \pmod 3$, so $\mathrm{err} = \Pr[N \equiv 2] \le \Pr[N \ge 2] \le \mathbb{E}\binom{N}{2} = \binom{m}{2}s^2$, by pairwise independence of the columns. Also $\Pr[h \ne u] = s$; the other nonzero constant differs from h off E, so with probability at least $1 - s \ge s$; and κw_i agrees with h with probability at most $\Pr[\kappa w_i = u] + \Pr[E] = \frac{1}{3} + s$, hence differs from h with probability at least $\frac{2}{3} - s \ge s$. □

So for polymorphisms of $P_{\not=}^{(m)}$ = under π_m the supremum of the possible distances at error at most δ (over all n) is $(\frac{2}{m} + O(\delta))\delta$ as $\delta \to 0$ if $3 \mid m$, as for $m = 3$ (Corollary 8.68), and $(1 + O(m\sqrt{\delta}))\sqrt{2\delta/(m(m-1))}$ if $3 \nmid m$: the upper bounds are (c), and by (d) the distance s occurs at error at most $\frac{m}{2}s$, respectively $\binom{m}{2}s^2$, for every multiple s of 3^{1-n} in $(0, \frac{1}{3}]$. When $3 \nmid m$ the constants 1 and 2 are polymorphisms, and near them the dependence is a square root, while near the functions κw_i it is linear, by the proof of (c). Part (a) is the analogue for $P_{\not=}^{(m)}$ = of the guarantee in the second question of [2, Section 9] (asked there for Boolean predicates), for all pairs at once. For $3 \mid m$, part (b) is known in substance: if $\mathbf{c}_1 = x$ and $\mathbf{c}_2 = \cdots = \mathbf{c}_m = y$, the rows lie in $P_{\not=}^{(m)}$ = iff $x_i \ne y_i$ for all i (as $m - 1 \equiv -1 \pmod 3$), and then a polymorphism h satisfies $h(x) \ne h(y)$; so h is a proper 3-colouring of the n-th direct power of K_3, hence of the form $\rho(x_i)$ by the theorem of Greenwell and Lovász [14] (Remark 8.74(c)), and every permutation ρ of $\mathbb{Z}_3$ is affine. The same argument gives Corollary 8.68(b).

Remark 8.76. The script `neqm_poly_check.py` (directory `numerics`) checks (b) for $m = 4,5,6,7$ and $n = 1,2$ by exhaustive search over all h ($2n+2$ polymorphisms for $m = 4,5,7$ and $6n$ for $m = 6$); the error formulas, the error bounds and the distance claim of (d) for $m = 4,5,6,7,9$ and $n = 2,3,4$, computing the law of N modulo 3 separately from the error; the two inequalities used in the proof of (c) for random h near polymorphisms ($m = 4,\dots,7$, $n = 2,3$; for $m = 6,7$ and $n = 2$ the first one is vacuous, and for $m = 6$ there is no second one); and the step behind (a), which does not depend on δ: for random tuples with prescribed equalities ($m = 4$, $n = 2,3$ and $m = 5$, $n = 2$), whenever some tuple of type (A_m) or (B_m) is within distance less than $\frac{1}{2}$, some such tuple with $g_j = g_l$ whenever $f_j = f_l$ is within the same distance.

Every full-support distribution for m summands. Let $m \ge 4$, let μ be a distribution on $P_{\not=}^{(m)}$ = with full support, $\alpha = \min\mu$ (so $\alpha \le \frac{1}{2\cdot 3^{m-1}} \le \frac{1}{54}$), and let μ_j be its marginals. As for $m = 3$, the rows of the matrix with columns $\mathbf{c}_1,\dots,\mathbf{c}_m$ are independent samples of μ, $\mathrm{err}_\mu(\mathbf{f})$ is the probability of a failure, $\mathrm{dist}(\mathbf{f},\mathbf{g}) = \max_j \Pr_{\mu_j^{\otimes n}}[f_j \ne g_j]$, $\beta_j(k)$ and dc refer to μ_j as before Lemma 8.60, and, as μ has full support, the tuples with error 0 are the generalized polymorphisms of Theorem 8.70(a). For distinct $a,b,c \in [m]$ let $R = [m] \setminus \{a,b,c\}$ and $T(y) = (y_a + \sum_{j\in R} y_j,\ y_b,\ y_c)$. As in Lemma 8.69(b), $y \mapsto (T(y), (y_j)_{j\in R})$ is a bijection of $P_{\not=}^{(m)}$ = onto $P_{\not=} \times \mathbb{Z}_3^R$; let μ_T and μ_R be the laws of $T(\mathbf{y})$ and of $(\mathbf{y}_j)_{j\in R}$ for $\mathbf{y} \sim \mu$. So μ_T is a distribution on $P_{\not=}$ whose second and third marginals are μ_b and μ_c, and $\min\mu_T \ge 3^{m-3}\alpha \ge 3\alpha$, as $\mu_T(p)$ is the sum of the masses of 3^{m-3} elements of $P_{\not=}^{(m)}$ =. By Theorem 8.66 the stability of μ_T is already known; this replaces the chain of Lemma 8.65, and since the decomposition can be made for every pair $\{b,c\}$, Step 5 of the proof of Theorem 8.61 is not needed.

Lemma 8.77. *Let $m \geq 4$, μ and α be as above, and let $\mathbf{f}$ be an m-tuple of functions on $\mathbb{Z}_3^n$ with $\mathrm{err}_\mu(\mathbf{f}) = \delta \leq \alpha^4/(4(m+1))$.*

(a) If $\beta_t(k) \geq \frac{1}{2}$ for some $t \in [m]$ and $k \in [n]$, there is a tuple $\mathbf{g}$ of type (B_m) with coordinate k such that $\Pr_{\mu_l^{\otimes n}}[f_l \neq g_l] \leq 2\delta/\alpha^4$ for every $l \in [m]$.

(b) Let $C' = C_{3\alpha}$ and $c' = c_{3\alpha}$ be the constants of Theorem 8.66 for 3α, let $r = r_\alpha$ (Lemma 8.35), $\delta_1 = (\alpha/(2C'))^{1/c'}$, $K = C' + \frac{15}{\alpha\delta_1}$ and $E_0 = \frac{4}{\alpha}K^{r/2}$. If $\beta_t(k) < \frac{1}{2}$ for all t and k, then $\min\big(\mathrm{dc}(f_b), \mathrm{dc}(f_c)\big) \leq E_0\,\delta^{c'r/2}$ for all $b \neq c$.

Proof. (a) is the extension of Lemma 8.60(a) to $P_{/}^{(m)} =$ proved in Remark 8.71(c).

(b) Fix $b \neq c$, choose $a \in [m]\backslash\{b,c\}$, and let R, T, μ_T and μ_R be as above. As $3\alpha \leq \min\mu_T \leq \frac{1}{18}$, C_x is decreasing and c_x is increasing in $x \in (0, \frac{1}{18}]$, Theorem 8.66 shows that μ_T is (C', c')-stable; also $C' \geq 1$ and $c' \leq 1$.

The decomposition. Identify a row y with $(T(y), (y_j)_{j\in R})$, let $\sigma = \frac{\alpha}{2}$ and $\nu = (\mu - \sigma\,\mu_T \otimes \mu_R)/(1-\sigma)$. As $\sigma\,\mu_T \otimes \mu_R \leq \frac{\alpha}{2} \leq \frac{\mu}{2}$, ν is a distribution on $P_{/}^{(m)} =$ with $\nu \geq \frac{\mu}{2} \geq \frac{\alpha}{2}$. Its b-th and c-th marginals are $\nu_b = (\mu_b - \sigma\mu_b)/(1-\sigma) = \mu_b$ and $\nu_c = \mu_c$, and, as every pair of values of y_b, y_c has $2\cdot 3^{m-3}$ completions in $P_{/}^{(m)} =$, the law ν_{bc} of these two entries of a ν-row satisfies $\nu_{bc}(u,v) \geq 3^{m-3}\alpha \geq \alpha\,\nu_b(u)\nu_c(v)$. Choose a random set $\mathbf{L} \subseteq [n]$ containing each i independently with probability σ; let the rows in $\mathbf{L}$ be independent samples of $\mu_T \otimes \mu_R$ and the rows outside $\mathbf{L}$ independent samples of ν, so that the rows are independent samples of μ. For a fixed L let ρ consist of the rows outside L and of the entries $y_{i,j}$ with $i \in L$ and $j \in R$; given L and ρ, the remaining randomness is $(T(\mathbf{y}_i))_{i\in L}$, a family of independent samples of μ_T. Put $s_i = \sum_{j\in R} y_{i,j}$ for $i \in L$, and let $\mathbf{x} \in \mathbb{Z}_3^L$ have the entries $y_{i,a} + s_i$, $i \in L$, i.e. the first entries of the $T(\mathbf{y}_i)$. The columns $\mathbf{c}_j$, $j \in R$, are determined by ρ, and $\mathbf{c}_a$ is determined by $\mathbf{x}$ and ρ; so $\sum_j f_j(\mathbf{c}_j) = F(\mathbf{x}) + f_b(\mathbf{c}_{b,L}, \rho_b) + f_c(\mathbf{c}_{c,L}, \rho_c)$ for a function F on $\mathbb{Z}_3^L$ depending on ρ, where ρ_t denotes the entries of the t-th column outside L and $\mathbf{c}_{t,L}$ those in L. Hence $\mathbf{f}^{L,\rho} = (F, f_b(\cdot,\rho_b), f_c(\cdot,\rho_c))$ is a triple of functions on $\mathbb{Z}_3^L$ whose error under μ_T, $\delta(L,\rho)$, is the conditional probability of a failure; let $\delta_L = \mathbb{E}_\rho\delta(L,\rho)$, so $\mathbb{E}_{\mathbf{L}}\delta_{\mathbf{L}} = \delta$. For $t \in \{b,c\}$ the entries of the t-th column have law μ_t both inside and outside L, so $\rho_t \sim \mu_t^{\otimes L^c}$.

Steps 1–4. We now run Steps 1–4 of the proof of Lemma 8.64 for the pair (b,c) in place of $(1,2)$ (after permuting the coordinates of $P_{/}=$, so that the second and third entries of $T(y)$ come first), with $\pi' = \mu_T$, $a = \alpha$ and $a' = 3\alpha$, so that $\theta = \min(2a', \frac{a}{2}) = \frac{\alpha}{2}$ and δ_1, K, r are as in the statement. These steps use only: the stability of π', applied to the triples $\mathbf{f}^{L,\rho}$; that, given L and ρ, the rows $T(\mathbf{y}_i)$, $i \in L$, are independent samples of μ_T, the pairs $(\rho_{b,i}, \rho_{c,i})$, $i \notin L$, are independent samples of ν_{bc}, and the b-th and the c-th column are independent of $\mathbf{L}$; the two types of generalized polymorphisms of $P_{/}=$ (Theorem 8.55(a)); that the free and the fixed entries of the b-th and the c-th column have the laws μ_b and μ_c; that $\min\mu_t \geq 6a$ for $t \in \{b,c\}$ (indeed $\mu_t(v) \geq 2\cdot 3^{m-2}\alpha$); that every value has probability at least $6a'$ under the corresponding marginal of π' (it is the entry of six elements of $P_{/}=$); the bound $\nu_{bc} \geq a\,\nu_b\nu_c$, for Lemma 8.59 with parameter a and Lemma 8.35 with $\kappa = a$; and $\beta_b(k), \beta_c(k) < \frac{1}{2}$ for all k. With $D_t(L) = \mathbb{E}_{\rho_t}\mathrm{dc}(f_t(\cdot,\rho_t))$, they give $\mathbb{E}_{\mathbf{L}}D_b(\mathbf{L})\cdot\mathbb{E}_{\mathbf{L}}D_c(\mathbf{L}) \leq K^r\delta^{c'r}$ and $\mathbb{E}_{\mathbf{L}}D_t(\mathbf{L}) \geq \frac{a}{4}\mathrm{dc}(f_t)$ for $t \in \{b,c\}$. Hence $\mathrm{dc}(f_b)\,\mathrm{dc}(f_c) \leq (\frac{4}{\alpha})^2K^r\delta^{c'r}$, which gives the claim. □

Theorem 8.78 (m summands, every full-support distribution)**.** *Let $m \geq 4$ and let μ be a distribution on $P_{/}^{(m)} =$ with full support and $\alpha = \min\mu$. For every n and every m-tuple $\mathbf{f}$ of functions on $\mathbb{Z}_3^n$ with $\mathrm{err}_\mu(\mathbf{f}) = \delta$ there is a generalized polymorphism $\mathbf{g}$, of type (A_m) or (B_m), with*

$$\mathrm{dist}(\mathbf{f},\mathbf{g}) \;\leq\; \exp\Big(1.22\cdot 10^{10}\,\alpha^{-16}\ln\frac{1}{\alpha}\Big)\,\delta^{\alpha^9/8000},$$

distances being measured under $\mu_j^{\otimes n}$.

Proof. Let c', r, E_0 be as in Lemma 8.77(b), $c = c'r/2$ and $C = \max\big(1+(m-1)E_0,\ 4(m+1)\alpha^{-4}\big)$; note $c \le 1$, so $\delta \le \delta^c$. If $\delta > \alpha^4/(4(m+1))$, every generalized polymorphism is within $1 < 4(m+1)\alpha^{-4}\delta \le C\delta^c$. If $\delta \le \alpha^4/(4(m+1))$ and $\beta_t(k) \ge \frac12$ for some t and k, Lemma 8.77(a) gives $\mathbf{g}$ of type (B_m) with $\operatorname{dist}(\mathbf{f},\mathbf{g}) \le 2\delta/\alpha^4 \le C\delta^c$. Otherwise, by Lemma 8.77(b), at most one index j has $\operatorname{dc}(f_j) > \varepsilon := E_0\delta^c$. Let j_0 be this index if there is one, and $j_0 = m$ otherwise; for $j \ne j_0$ let u_j be the most likely value of f_j, so $\Pr[f_j(\mathbf{c}_j) \ne u_j] \le \varepsilon$, and let $U = \sum_{j\ne j_0} u_j$. The event that $f_j(\mathbf{c}_j) = u_j$ for all $j \ne j_0$ and $f_{j_0}(\mathbf{c}_{j_0}) = -U$ is a failure, so $\Pr[f_{j_0}(\mathbf{c}_{j_0}) = -U] \le \delta + (m-1)\varepsilon$. Let $g_j \equiv u_j$ for $j \ne j_0$, and $g_{j_0} = f_{j_0}$ except that $g_{j_0} = 1 - U$ where $f_{j_0} = -U$. Then $\mathbf{g}$ is of type (A_m) and $\operatorname{dist}(\mathbf{f},\mathbf{g}) \le \delta + (m-1)\varepsilon \le C\delta^c$.

For the constants, let $x = 1/\alpha \ge 54$ and $\ell = \ln x$. As $r_\alpha \ge \frac{\alpha}{2}$ and $c' = \frac{3^8\alpha^8}{13107200}$, $c \ge \frac{6561}{52428800}\alpha^9 > \frac{\alpha^9}{8000}$, so $\delta^c \le \delta^{\alpha^9/8000}$. Next, $\ln C' = 1.2\cdot 10^{11}(3\alpha)^{-9}\ln\frac{1}{3\alpha} \le 6.097\cdot 10^6 x^9\ell$ and $1/c' \le 1997.8\,x^8$, so $\ln\frac{1}{\delta_1} = (\ln C' + \ln(2x))/c' \le 1.2181\cdot 10^{10}x^{17}\ell$. As $\delta_1 \le \frac{\alpha}{2C'}$, $K \le (1+15x)/\delta_1$, and as $r_\alpha \le 2\alpha$, $\ln E_0 \le \ln(4x) + \alpha\big(\ln(16x) + \ln\frac{1}{\delta_1}\big) \le \ln(4x) + \alpha\ln(16x) + 1.2181\cdot 10^{10}x^{16}\ell$. Finally $E_0 \ge 1$ and $m \le 1 + \log_3\frac{x}{2} \le x$ give $\ln C \le \max\big(\ln m + \ln E_0,\ \ln(4(m+1)) + 4\ell\big) \le 1.22\cdot 10^{10}x^{16}\ell$. □

Corollary 8.79 (Equal functions and polymorphisms, every full-support distribution)**.** *Let $m \ge 4$, let μ be a distribution on $P_{/}^{(m)}$ = with full support and $\alpha = \min\mu$, let $C_* = \exp(1.22\cdot 10^{10}\alpha^{-16}\ln\frac{1}{\alpha})$ and $c_* = \alpha^9/8000$ be the constant and the exponent of Theorem 8.78, and let $n \ge 1$.*

(a) For every $\mathbf{f}$ with $\operatorname{err}_\mu(\mathbf{f}) = \delta$ there is a generalized polymorphism $\mathbf{g}$ of type (A_m) or (B_m) with $\operatorname{dist}(\mathbf{f},\mathbf{g}) \le 2C_*\delta^{c_*}$ such that $g_j = g_l$ whenever $f_j = f_l$ and $\mu_j = \mu_l$.

(b) Let $3 \mid m$, $\mu_1 = \cdots = \mu_m =: \nu$, and $\delta = \operatorname{err}_\mu(h,\dots,h)$ for a function $h\colon \mathbb{Z}_3^n \to \mathbb{Z}_3$. Then $\Pr_{\nu^{\otimes n}}[h \ne \kappa w_i + d] \le K_*\delta$ for some i, κ and d, where $K_* = (2C_*/\alpha)^{1/c_*}$; the functions $\kappa w_i + d$ are the polymorphisms, by Corollary 8.75(b), which concerns only the support.

Proof. (a) Let $\mathbf{g}$ be given by Theorem 8.78 and $\eta = \operatorname{dist}(\mathbf{f},\mathbf{g}) \le C_*\delta^{c_*}$. If $\eta \ge \frac12$, the tuple $g_1 = \cdots = g_m = w_1$, which is of type (B_m) for every m (all $d_j = 0$), is within $1 \le 2C_*\delta^{c_*}$. Let $\eta < \frac12$ and call j and l equivalent if $f_j = f_l$ and $\mu_j = \mu_l$. For equivalent j, l we have $\Pr_{\mu_j^{\otimes n}}[g_j \ne g_l] \le 2\eta < 1$, and the proof of Corollary 8.75(a) applies verbatim (it uses only this inequality, the forms of the types (A_m) and (B_m), and $\Pr_{\mu_b^{\otimes n}}[f_b \ne u_l] = \Pr_{\mu_l^{\otimes n}}[f_l \ne u_l]$ for equivalent b, l): it gives a tuple of type (A_m) or (B_m) within η in which equivalent indices carry equal functions.

(b) As $C_* \ge 1$, $c_* \le \frac15$ and $m \le \frac{1}{\alpha}$, $K_* \ge (2/\alpha)^5 \ge 4(m+1)\alpha^{-4}$; so $K_*\delta \ge 1$ if $2C_*\delta^{c_*} > \alpha$ or $\delta > \alpha^4/(4(m+1))$, and then every $\kappa w_i + d$ will do. Otherwise (a), applied to $\mathbf{f} = (h,\dots,h)$, gives a generalized polymorphism $(g,\dots,g)$ with $\Pr_\nu[h \ne g] \le 2C_*\delta^{c_*} \le \alpha$. As $3 \mid m$, g is not constant, so $g = \kappa w_i + d$, and Lemma 8.60(b) for $P_{/}^{(m)}$ = (Remark 8.71(c)) gives $\beta_1(i) \ge \frac12$. By Lemma 8.77(a) some $\mathbf{g}'$ of type (B_m) satisfies $\Pr_\nu[h \ne g_l'] \le 2\delta/\alpha^4 \le \frac{1}{2(m+1)}$ for every l; then $\Pr_\nu[g_j' \ne g_l'] < 1$, so all g_l' are equal to some $\kappa' w_i + d'$, and $\Pr_\nu[h \ne \kappa' w_i + d'] \le 2\delta/\alpha^4 \le K_*\delta$. □

So the first (qualitative) question of [2, Section 9] has a positive answer, with polynomial dependence, for every $P_{/}^{(m)}$ = under every full-support distribution, and by Corollary 8.73 for every largest predicate over $\mathbb{Z}_3$ without a flexible coordinate ($m \ge 3$; Theorem 8.66 for $m = 3$). Their question about polynomial dependence concerns the Boolean setting of [2, Theorem 1.6]; and the stronger linear form fails here, since the exponent $\frac12$ is sharp under the uniform distribution (Theorem 8.55 for $m = 3$). For permutations, Filmus [9, Theorem 3.9] gives an $\varepsilon^{1/4}$ bound for Perm_n under the uniform distribution. By Remark 8.71(c) the exponent must depend on μ. Corollary 8.79(a) is

the guarantee of their second question (asked there for Boolean predicates) for $P^{(m)}_{/}=$, as Corollary 8.68(a) is for $m=3$. As for $m=3$, the exponent and the constant are far from optimal; under π_m, Theorem 8.70 gives the exponent $\frac{1}{2}$. The script `neqm_mu_check.py` (directory `numerics`) checks, for random μ with $m=4,5$ (generic, nearly degenerate, and with very unequal masses), every pair $\{b,c\}$ and every a, the facts about T, μ_T, μ_R and ν used in the proof of Lemma 8.77(b); for $m=4$ and $n=2$ the identity $\mathbb{E}_{\mathbf{L}}\delta_{\mathbf{L}}=\delta$, with $\delta(L,\rho)$ computed as the error of $\mathbf{f}^{L,\rho}$ under μ_T, by exact enumeration; the counts, the kernel bound and the final inequality used in Remark 8.71(c); and the constants of Theorem 8.78 on a grid of values $\alpha\in[10^{-12},\frac{1}{54}]$, for every admissible m.

Uniform marginals for m summands. For $m=3$, Theorem 8.61(a) and Proposition 8.62(b) show that the best exponent for full-support distributions with uniform marginals and smallest atom α is of order α. The same holds for every m, with α replaced by $\alpha'=3^{m-3}\alpha$, which ranges over $(0,\frac{1}{18}]$ for every m ($\alpha'=\frac{1}{18}$ only for π_m). For $\alpha'>\frac{1}{45}$, where Proposition 8.62(b) and Theorem 8.80(b) do not apply, the trivial bound $c\le 1<45\alpha'$ gives the upper bound of order α': under every μ with full support and uniform marginals, the tuple of Theorem 8.55(c), resp. Theorem 8.70(c), with $\Pr[E]=\eta$ has error at most η (a failure needs the last column, which is uniform, to lie in E) and distance at least η (the proof of the distance bound uses only part (a), which concerns only the support, and the uniformity of the columns). Proposition 8.82 below gives the upper bound $9\alpha'$ for every $\alpha'\le\frac{1}{18}$, sharp up to a factor $\frac{9}{8}$.

Theorem 8.80 (m summands, uniform marginals)**.** *Let $m\ge 4$, let μ be a distribution on $P^{(m)}_{/}=$ with full support and uniform marginals, $\alpha=\min\mu$ and $\alpha'=3^{m-3}\alpha$, let $r=r_{9\alpha'}$ (Lemma 8.35), so $r\ge 8\alpha'$, and let δ_m be as in Theorem 8.70.*

(a) For every $\mathbf{f}$ with $\mathrm{err}_\mu(\mathbf{f})=\delta$ there is a generalized polymorphism $\mathbf{g}$ of type (A_m) or (B_m) with $\mathrm{dist}(\mathbf{f},\mathbf{g})\le C\delta^{r/2}\le C\delta^{4\alpha'}$, where $C=\max\left((4(m+1))^{1/4}\alpha^{-1},\ \frac{m}{2}\alpha'^{-3/2}\delta_m^{-1/2}\right)$.

(b) Let $\alpha>0$ be such that $\alpha'=3^{m-3}\alpha\le\frac{1}{45}$, and let $\mu^{\mathrm{sh}}_{m,\alpha}$ give mass α to each of the $4\cdot 3^{m-2}$ elements of $P^{(m)}_{/}=$ with $a_m\ne a_{m-1}+1$ and mass $\frac{1-4\cdot 3^{m-2}\alpha}{2\cdot 3^{m-2}}$ to each of the other $2\cdot 3^{m-2}$ elements. Then $\mu^{\mathrm{sh}}_{m,\alpha}$ has full support, minimum α and uniform marginals, and if constants $C,c>0$ are such that every $\mathbf{f}$ (for every n) is within $C\,\mathrm{err}_{\mu^{\mathrm{sh}}_{m,\alpha}}(\mathbf{f})^c$ of a generalized polymorphism, then $c\le\frac{9}{2}e^2\alpha'$.

Proof. (a) We follow the proof of Theorem 8.61(a). Note $r<9\alpha'\le\frac{1}{2}$, so $\delta\le\delta^{r/2}$ and $\alpha^{-2r}\le\alpha^{-1}$. If $\delta>\alpha^4/(4(m+1))$, every generalized polymorphism is within $1<(4(m+1)\delta/\alpha^4)^{r/2}\le(4(m+1))^{1/4}\alpha^{-1}\delta^{r/2}$. If $\delta\le\alpha^4/(4(m+1))$ and $\beta_t(k)\ge\frac{1}{2}$ for some t and k, Lemma 8.77(a) gives $\mathbf{g}$ of type (B_m) within $2\delta/\alpha^4=2\delta^{r/2}\delta^{1-r/2}\alpha^{-4}\le 2\alpha^{-2r}\delta^{r/2}\le(4(m+1))^{1/4}\alpha^{-1}\delta^{r/2}$, as $\delta^{1-r/2}\le\alpha^{4-2r}$ and $(4(m+1))^{1/4}\ge 2$. So we may assume that $\beta_t(k)<\frac{1}{2}$ for all t and k.

The decomposition. Let $\sigma=9\alpha'=3^{m-1}\alpha\le\frac{1}{2}$ and $\nu=(\mu-\sigma\pi_m)/(1-\sigma)$. As $\pi_m\equiv\frac{1}{2\cdot 3^{m-1}}$ on $P^{(m)}_{/}=$, $\sigma\pi_m\equiv\frac{\alpha}{2}$, so $\nu\ge\frac{\alpha}{2}$ on $P^{(m)}_{/}=$; the marginals of ν are uniform; and, as every pair of entries has $2\cdot 3^{m-3}$ completions, the law ν_{jl} of two entries of a ν-row satisfies $\nu_{jl}(a,b)\ge 3^{m-3}\alpha=\alpha'=9\alpha'\,\nu_j(a)\nu_l(b)$. As in the proof of Theorem 8.61, choose $\mathbf{L}\subseteq[n]$ containing each i independently with probability σ, let the rows in $\mathbf{L}$ be independent samples of π_m and the others independent samples of ν, and for fixed L let $\rho=(\rho_1,\dots,\rho_m)$ be the columns of the rows outside L and $\mathbf{f}^{L,\rho}=(f_j(\cdot,\rho_j))_{j\in[m]}$, an instance of Theorem 8.70 with $|L|$ coordinates (for $L=\emptyset$ its functions are constants, which form a tuple of type (A_m) if $\delta(L,\rho)<1$), with error $\delta(L,\rho)$; $\delta_L=\mathbb{E}_\rho\delta(L,\rho)$ and $\mathbb{E}_{\mathbf{L}}\delta_{\mathbf{L}}=\delta$; each ρ_j is uniform on $\mathbb{Z}_3^{L^c}$. By Theorem 8.70(b), as $2m\delta_m^{1/4}<1$, if $\delta(L,\rho)\le\delta_m$ then some $\mathbf{g}'$ of type (A_m) or (B_m) on $\mathbb{Z}_3^L$ is within $2\sqrt{\delta(L,\rho)}<\frac{1}{10}$ of $\mathbf{f}^{L,\rho}$.

Step 1. Define $P_j(\rho_j) \in L \cup \{\perp\}$ as in Step 1 of the proof of Theorem 8.61. If $P_1(\rho_1), \dots, P_m(\rho_m)$ are not all equal, then $\delta(L,\rho) > \delta_m$: a tuple $\mathbf{g}'$ of type (B_m) with coordinate i within $\frac{1}{10}$ would give $P_j(\rho_j) = i$ for all j, and one of type (A_m), in which $m-1$ functions are constants and the last takes at most two values, would give $P_j(\rho_j) = \perp$ for all j. By Markov's inequality $\Pr_\rho[P_1(\rho_1) \neq P_2(\rho_2)] \le \delta_L/\delta_m$, and, as there, Lemma 8.59 for the first two entries of a ν-row with $\theta = 9\alpha'$ (the kernel satisfies $K(a,a') = 9\sum_b \nu_{12}(a,b)\nu_{12}(a',b) \ge 9\alpha' \cdot \frac{1}{3}$) gives $\mathrm{dc}(P_1(\rho_1)) \le \eta_L := 2\delta_L/(9\alpha'\delta_m)$. If $\eta_L \le \frac{1}{7}$ and the most likely value of $P_1(\rho_1)$ were some $i \in L$, then $\beta_1(i) \ge \frac{7}{10}(1-\eta_L) \ge \frac{3}{5}$, as there; so $\Pr_\rho[P_1(\rho_1) \neq \perp] \le 7\eta_L$ in all cases.

Step 2. If $P_1(\rho_1) = \perp$ and $\delta(L,\rho) \le \delta_m$, then $\mathbf{g}'$ is of type (A_m): for some index s, $g'_t \equiv u_t$ for $t \neq s$, and g'_s never takes the value $-U$, where $U = \sum_{t\neq s} u_t$. Write $h_t = f_t(\cdot,\rho_t)$, $\varepsilon_0 = 2\sqrt{\delta(L,\rho)}$ and $p_t = \Pr[h_t \neq u_t] \le \varepsilon_0$ for $t \neq s$; then $\mathrm{dc}(h_t)\,\mathrm{dc}(h_{t'}) \le p_t p_{t'} \le 4\,\delta(L,\rho)$ for $t,t' \neq s$. Given L and ρ, the rows in L are independent samples of π_m, and every pair of entries has $2\cdot 3^{m-3}$ completions in $P^{(m)}_{/\!=}$, so any two of the columns restricted to L are independent and uniform. Fix $t \neq s$ and, for $a \neq u_t$, let $q_a = \Pr[h_s = u_t - a - U]$. As in Step 2 of the proof of Theorem 8.61, the disjoint failures where $h_t = a$, $h_{t'} = u_{t'}$ for all $t' \notin \{s,t\}$ and $h_s = u_t - a - U$, $a \neq u_t$, give $\delta(L,\rho) \ge p_t\big(\min_{a\neq u_t} q_a - \sum_{t'\notin\{s,t\}} p_{t'}\big)$; the values $u_t - a - U$, $a \neq u_t$, are the two values other than $-U$, so $\mathrm{dc}(h_s) \le \Pr[h_s = -U] + \min_{a\neq u_t} q_a \le \varepsilon_0 + \min_{a\neq u_t} q_a$; hence $\mathrm{dc}(h_t)\,\mathrm{dc}(h_s) \le \delta(L,\rho) + p_t\big((m-2)\varepsilon_0 + \varepsilon_0\big) \le (4m-3)\,\delta(L,\rho)$. So $m_{jl}(L,\rho) := \mathrm{dc}(f_j(\cdot,\rho_j))\,\mathrm{dc}(f_l(\cdot,\rho_l)) \le (4m-3)\,\delta(L,\rho)$ for every pair $j \neq l$ in this case, $m_{jl} \le 1$ always, and we get, as there,

$$\mathbb{E}_\rho\, m_{jl}(L,\rho) \le (4m-3)\,\delta_L + 7\eta_L + \frac{\delta_L}{\delta_m} \le K_1\delta_L, \qquad K_1 := \frac{2}{\alpha'\delta_m},$$

since $4m-3+\frac{14}{9\alpha'\delta_m} + \frac{1}{\delta_m} = \frac{(4m-3)\alpha'\delta_m + 14/9 + \alpha'}{\alpha'\delta_m} \le \frac{2}{\alpha'\delta_m}$ (as $(4m-3)\alpha'\delta_m + \alpha' \le \frac{4}{9}$).

Steps 3 and 4. For every pair $j \neq l$ (all marginals are uniform), Step 3 of the proof of Theorem 8.61, with Lemma 8.35 for $\kappa = 9\alpha'$, gives $\mathbb{E}_{\mathbf{L}} D_j(\mathbf{L}) \cdot \mathbb{E}_{\mathbf{L}} D_l(\mathbf{L}) \le K_1^r \delta^r$, and Step 4 gives $\mathbb{E}_{\mathbf{L}} D_j(\mathbf{L}) \ge \frac{\sigma}{3}\mathrm{dc}(f_j) = 3\alpha'\,\mathrm{dc}(f_j)$.

Conclusion. So $\mathrm{dc}(f_j)\,\mathrm{dc}(f_l) \le K_1^r\delta^r/(9\alpha'^2)$ for all $j \neq l$, and at most one index j_0 has $\mathrm{dc}(f_{j_0}) > \varepsilon := K_1^{r/2}\delta^{r/2}/(3\alpha')$; if there is none, let $j_0 = m$. As in the proof of Theorem 8.78, with u_j the most likely value of f_j $(j \neq j_0)$ and $U = \sum_{j\neq j_0} u_j$, we get $\Pr[f_{j_0} = -U] \le \delta + (m-1)\varepsilon$ and a tuple $\mathbf{g}$ of type (A_m) with $\mathrm{dist}(\mathbf{f},\mathbf{g}) \le \delta + (m-1)\varepsilon$. As $K_1 \ge 1$ and $r \le 1$, $\varepsilon \le \frac{\sqrt{2}}{3}\alpha'^{-3/2}\delta_m^{-1/2}\delta^{r/2}$; with $X = \alpha'^{-3/2}\delta_m^{-1/2} \ge 18^{3/2}\cdot 4\cdot 10^7 > 3\cdot 10^9$, $\delta + (m-1)\varepsilon \le \big(1 + \frac{\sqrt{2}}{3}(m-1)X\big)\delta^{r/2} \le \frac{m}{2}X\delta^{r/2}$.

(b) An element $(a_1,\dots,a_{m-2},b,b+1)$ is in $P^{(m)}_{/\!=}$ iff $b \neq \sum_{j\le m-2} a_j + 1$, so there are $2\cdot 3^{m-2}$ elements with $a_m = a_{m-1}+1$, and $\frac{1-4\cdot 3^{m-2}\alpha}{2\cdot 3^{m-2}} \ge \alpha$ as $\alpha \le \frac{1}{2\cdot 3^{m-1}}$. The translations $a \mapsto a + t$ with $\sum_j t_j = 0$ and $t_{m-1} = t_m$ preserve $P^{(m)}_{/\!=}$ and the condition $a_m = a_{m-1}+1$, hence $\mu^{\mathrm{sh}}_{m,\alpha}$, and they shift each entry by every element of $\mathbb{Z}_3$ (Remark 8.71(c)); so the marginals are uniform. Let n, $\tau = 2e^2\alpha' < \frac{1}{3}$, ℓ, Y, Z be as in the proof of Proposition 8.62(b), and $f_1 = \cdots = f_{m-2} \equiv 0$, $f_{m-1} = \mathbf{1}_Y$, $f_m = 1 + \mathbf{1}_Z$. A failure means $\mathbf{c}_{m-1} \in Y$ and $\mathbf{c}_m \in Z$. A row has $y_{m-1} = 0$ and $y_m \neq 1$ with probability $4\cdot 3^{m-3}\alpha = 4\alpha'$: for each $y_m \in \{0,2\}$, the first $m-2$ entries are any of the $2\cdot 3^{m-3}$ tuples whose sum is not $-y_m$, and all these elements have mass α. The proof of Proposition 8.62(b) now applies verbatim, with $\mathbf{c}_{m-1}, \mathbf{c}_m$ in place of $\mathbf{y},\mathbf{z}$ and $4\alpha'$ in place of 4α; for the distance, a tuple of type (B_m) has $\Pr[f_1 \neq g_1] = \frac{2}{3}$, and in a tuple of type (A_m) g_{m-1} or g_m is a constant. Hence $c \le \frac{9}{4}\tau = \frac{9}{2}e^2\alpha'$. □

So for every $m \ge 3$, the supremum of the exponents c such that every full-support μ on $P^{(m)}_{/\!=}$ with uniform marginals and $\min\mu = \alpha$ admits a bound $\mathrm{dist} \le C_\mu\,\mathrm{err}_\mu^c$ is of order $\alpha' = 3^{m-3}\alpha$: it is at least $r_{9\alpha'}/2 \ge 4\alpha'$ by Theorems 8.61(a) and 8.80(a), with explicit constants, and at most $\frac{9}{2}e^2\alpha'$ when

$\alpha' \le \frac{1}{45}$. Propositions 8.81 and 8.82 below sharpen this, for $\alpha' < \frac{1}{18}$, to the interval $[r_{18\alpha'}/2, 9\alpha']$, whose ends are within a factor $\frac{9}{8}$, uniformly in m. As $\mu^{\mathrm{sh}}_{m,\alpha}$ has full support, the exponent of Theorem 8.78 can be improved at most to $9 \cdot 3^{m-3}\alpha$. The script `neqm_unif_check.py` (directory `numerics`) checks, for $m = 4, 5$, the decomposition and the bounds $\nu \ge \frac{\alpha}{2}$, $\nu_{jl} \ge \alpha' = 9\alpha'\nu_j\nu_l$ and $K(a, a') \ge 9\alpha'\nu_j(a')$ for random distributions with uniform marginals; the distribution $\mu^{\mathrm{sh}}_{m,\alpha}$ exactly for $m = 4, 5, 6$, including that the four classes of rows given by $\mathbf{1}[y_{m-1} = 0]$ and $\mathbf{1}[y_m = 1]$ have the same probabilities as for $\mu^{\mathrm{sh}}_{\alpha'}$ and $m = 3$; for $m = 4$, by brute force for $n = 3$, the inequality $N_0(\mathbf{c}_3) - N_1(\mathbf{c}_4) \le \mathbf{D}$, the law of $(N_0(\mathbf{c}_3), N_1(\mathbf{c}_4), \mathbf{D})$ and the error; the error bound exactly for $n \le 24$; the distance bound against all generalized polymorphisms for $n = 3, 6, 9$; and the numerical inequalities used for the constants.

A larger parameter for reverse hypercontractivity. In the proofs of Theorems 8.61 and 8.80 the weight of π_m in the decomposition is $\sigma = 9\alpha'$ (for $m = 3$, $\pi_3 = \pi$ and $\alpha' = \alpha$). A smaller weight leaves the ν-rows closer to μ, whose pair laws satisfy $\mu_{jl}(a, b) \ge 2 \cdot 3^{m-3}\alpha = 18\alpha'\mu_j(a)\mu_l(b)$ when μ_j and μ_l are uniform, so that reverse hypercontractivity can be applied with a parameter close to $18\alpha'$, at the cost of the constant. Similarly, Step 5 can use $\kappa = 9\alpha/(1 - 9\alpha)$ in place of 2α when one of μ_l, μ_m is uniform.

Proposition 8.81 (A larger parameter)**.** *Let $m \ge 3$, let μ be a distribution on $P^{(m)}_{/} =$ with full support, $\alpha = \min \mu$ and $\alpha' = 3^{m-3}\alpha$, and suppose $\mu \ne \pi_m$, i.e. $\alpha' < \frac{1}{18}$. For $0 < s \le \frac{1}{2}$ let*

$$\kappa_s = \frac{18\alpha'(1-s)}{1 - 18\alpha' s} \qquad \text{and} \qquad r = r_{\kappa_s} \quad \text{(Lemma 8.35)},$$

and let $\mathbf{f}$ be a tuple of functions on $\mathbb{Z}_3^n$ with $\mathrm{err}_\mu(\mathbf{f}) = \delta$.

(a) If all marginals of μ are uniform, there is a generalized polymorphism $\mathbf{g}$ with $\mathrm{dist}(\mathbf{f}, \mathbf{g}) \le C_s\delta^{r/2}$, where $C_s = \max\left((4(m+1))^{1/2}\alpha^{-2},\ s^{-1}M_m\right)$, with $M_3 = 1.5 \cdot 10^7\alpha^{-3/2}$ and $M_m = \frac{m}{4}\alpha'^{-3/2}\delta_m^{-1/2}$ for $m \ge 4$ (δ_m as in Theorem 8.70).

(b) If $m = 3$ and μ_1 and μ_2 are uniform, there is a generalized polymorphism $\mathbf{g}$ with $\mathrm{dist}(\mathbf{f}, \mathbf{g}) \le C_s\delta^{rr''/4}$, where $r'' = r_{\kappa''} \ge 8\alpha$ with $\kappa'' = 9\alpha/(1 - 9\alpha)$.

Proof. We follow the proofs of Theorem 8.61 ($m = 3$) and Theorem 8.80(a) ($m \ge 4$), with $\sigma = 18\alpha' s \le 9\alpha' \le \frac{1}{2}$ in place of $9\alpha'$. We have $\kappa_s \ge 18\alpha'(1 - s) \ge 9\alpha'$, and $\kappa_s < 1$ as $18\alpha' < 1$; so $\frac{8}{9}\kappa_s \le r < \kappa_s < 1$ by Lemma 8.35, and $r \ge 8\alpha'$. Also $\kappa'' < 1$ as $\alpha < \frac{1}{18}$ in case (b), so $8\alpha \le \frac{8}{9}\kappa'' \le r'' < 1$. We may assume $\delta > 0$: if $\delta = 0$ then, as μ has full support, $\mathbf{f}$ is itself a generalized polymorphism. (The strict inequalities between multiples of powers of δ below, and those in the conclusion of the proof of Theorem 8.61, are meant for $\delta > 0$; for $\delta = 0$ the same proofs give dist $= 0$.)

The opening. If $\delta > \alpha^4/(4(m+1))$, every generalized polymorphism is within $1 < (4(m+1)\delta/\alpha^4)^{r/2} \le (4(m+1))^{1/2}\alpha^{-2}\delta^{r/2}$. If $\delta \le \alpha^4/(4(m+1))$ and $\beta_t(k) \ge \frac{1}{2}$ for some t and k, then Lemma 8.60(a) ($m = 3$, where $4(m+1) = 16$) or Lemma 8.77(a) ($m \ge 4$) gives $\mathbf{g}$ within $3\delta/\alpha^4 = 3\delta^{r/2}\delta^{1-r/2}\alpha^{-4} \le 3\alpha^{-2r}\delta^{r/2} \le 3\alpha^{-2}\delta^{r/2}$, as $\delta^{1-r/2} \le \alpha^{4-2r}$. Both bounds are at most $C_s\delta^{r/2} \le C_s\delta^{rr''/4}$. So we may assume that $\beta_t(k) < \frac{1}{2}$ for all t and k.

The decomposition. Let $\nu = (\mu - \sigma\pi_m)/(1 - \sigma)$. As $\sigma\pi_m \equiv \sigma/(2 \cdot 3^{m-1}) = \alpha s$ on $P^{(m)}_{/} =$, we have $\nu \ge \alpha(1 - s)/(1 - \sigma) \ge \frac{\alpha}{2}$ there; the marginal ν_j is uniform whenever μ_j is; and, as every pair of entries has $2 \cdot 3^{m-3}$ completions, $\nu_{jl}(a, b) \ge 2\alpha'(1 - s)/(1 - \sigma) = \kappa_s/9$. So $\nu_{jl}(a, b) \ge \kappa_s\,\nu_j(a)\nu_l(b)$ when ν_j and ν_l are uniform, and $\nu_{jl}(a, b) \ge \alpha'$ always. Let $\mathbf{L}$, the rows, ρ, $\mathbf{f}^{L,\rho}$, $\delta(L, \rho)$ and δ_L be as in those proofs, now with each i in $\mathbf{L}$ independently with probability σ.

Steps 1 and 2 hold verbatim, with the same η_L and K_1: besides the π_m-rows in L and the uniformity of ν_1 and ν_2, they use only $\nu_{12}(a,b) \geq \alpha'$, which gives $\theta = 9\alpha'$ in Lemma 8.59. So $\mathbb{E}_\rho\, m_{jl}(L,\rho) \leq K_1\delta_L$ for all $j \neq l$, with $K_1 = 2\cdot 10^{15}/\alpha$ for $m = 3$ and $K_1 = 2/(\alpha'\delta_m)$ for $m \geq 4$. *Step 3* holds for every pair $j \neq l$ with μ_j and μ_l uniform, with Lemma 8.35 applied for $\kappa = \kappa_s$ to the pairs $(\rho_{j,i}, \rho_{l,i})$: $\mathbb{E}_{\mathbf{L}} D_j(\mathbf{L}) \cdot \mathbb{E}_{\mathbf{L}} D_l(\mathbf{L}) \leq K_1^r\delta^r$. *Step 4* gives $\mathbb{E}_{\mathbf{L}} D_j(\mathbf{L}) \geq \frac{\sigma}{3}\mathrm{dc}(f_j) = 6\alpha' s\,\mathrm{dc}(f_j)$ when μ_j is uniform. Hence $\mathrm{dc}(f_j)\,\mathrm{dc}(f_l) \leq K_1^r\delta^r/(36\alpha'^2 s^2)$ for these pairs. Let $\varepsilon = K_1^{r/2}\delta^{r/2}/(6\alpha' s)$.

Conclusion for $m = 3$. As $K_1 \geq 1$ and $r < 1$, $\varepsilon \leq \frac{\sqrt{2}}{6}\,10^{7.5} s^{-1}\alpha^{-3/2}\delta^{r/2} < 7.4536\cdot 10^6 s^{-1}\alpha^{-3/2}\delta^{r/2}$; and $\delta \leq \delta^{r/2} \leq 18^{-3/2}\alpha^{-3/2}\delta^{r/2} < 0.007\, s^{-1}\alpha^{-3/2}\delta^{r/2}$, as $\alpha \leq \frac{1}{18}$ and $s \leq \frac{1}{2}$. In case (a), two indices have $\mathrm{dc} \leq \varepsilon$, and the construction at the end of the proof of Theorem 8.61 gives $\mathbf{g}$ of type (A) with $\mathrm{dist}(\mathbf{f},\mathbf{g}) \leq \delta + 2\varepsilon < 1.5\cdot 10^7 s^{-1}\alpha^{-3/2}\delta^{r/2}$. In case (b), Steps 3 and 4 for the pair $\{1,2\}$ give one $j \in \{1,2\}$ with $\mathrm{dc}(f_j) \leq \varepsilon$. In Step 5 with $\varepsilon' = \varepsilon$, write $\{l,l'\} = [3]\backslash\{j\}$ with $l \in \{1,2\}$ and $l' = 3$ for the two indices called l, m there, and $v \in \mathbb{Z}_3$ for the index called s there. Then μ_l is uniform, and every entry of the law $\mu_{ll'}$ of the pairs $(\mathbf{c}_{l,i}, \mathbf{c}_{l',i})$ is at least 2α, as every pair of entries has two completions in $P_{\neq}$. So for $a, b \in \mathbb{Z}_3$ and $a' \neq a$, $\mu_{ll'}(a',b) = \frac{1}{3} - \sum_{b'\neq b}\mu_{ll'}(a',b') \leq \frac{1}{3} - 4\alpha$; hence $\mu_{l'}(b) \leq \mu_{ll'}(a,b) + \frac{2}{3} - 8\alpha$, and as $x \mapsto 3x/(x + \frac{2}{3} - 8\alpha)$ is increasing on $x > 0$ and $\mu_{ll'}(a,b) \geq 2\alpha$, we get $\mu_{ll'}(a,b) \geq \frac{6\alpha}{\frac{2}{3}-6\alpha}\mu_l(a)\mu_{l'}(b) = \kappa''\mu_l(a)\mu_{l'}(b)$. Lemma 8.35 with $\kappa = \kappa''$ gives $p_v q_v \leq (\delta+\varepsilon)^{r''}$ for every $v \in \mathbb{Z}_3$, and the rest of Step 5 gives a second index with $\mathrm{dc} \leq 2(\delta+\varepsilon)^{r''/2}$. With $Y = 7.46\cdot 10^6 s^{-1}\alpha^{-3/2} \geq 1$ we have $\delta + \varepsilon \leq Y\delta^{r/2}$, so $2(\delta+\varepsilon)^{r''/2} \leq 2Y^{1/2}\delta^{rr''/4} < 5463\, s^{-1/2}\alpha^{-3/4}\delta^{rr''/4} < 626\, s^{-1}\alpha^{-3/2}\delta^{rr''/4}$, as $\alpha^{-3/4} \leq 18^{-3/4}\alpha^{-3/2}$ and $s \leq 1$. As $\delta^{r/2} \leq \delta^{rr''/4}$, two indices have $\mathrm{dc} \leq \varepsilon'' := \max(\varepsilon, 2(\delta+\varepsilon)^{r''/2}) < 7.4536\cdot 10^6 s^{-1}\alpha^{-3/2}\delta^{rr''/4}$, and the same construction gives $\mathrm{dist}(\mathbf{f},\mathbf{g}) \leq \delta + 2\varepsilon'' < 1.5\cdot 10^7 s^{-1}\alpha^{-3/2}\delta^{rr''/4}$.

Conclusion for $m \geq 4$. At most one index j_0 has $\mathrm{dc}(f_{j_0}) > \varepsilon$, and the construction at the end of the proof of Theorem 8.80(a) gives $\mathbf{g}$ of type (A_m) with $\mathrm{dist}(\mathbf{f},\mathbf{g}) \leq \delta + (m-1)\varepsilon$. With $X = \alpha'^{-3/2}\delta_m^{-1/2} > 3\cdot 10^9$, as there, $\varepsilon \leq \frac{\sqrt{2}}{6} s^{-1} X\delta^{r/2}$, so $\delta + (m-1)\varepsilon \leq \big(1 + \frac{\sqrt{2}}{6}(m-1)s^{-1}X\big)\delta^{r/2} \leq \frac{m}{4} s^{-1} X\delta^{r/2}$, since $\frac{m}{4} - \frac{\sqrt{2}}{6}(m-1) \geq \frac{\sqrt{2}}{6} > \frac{1}{s^{-1}X}$. □

As $s \to 0$, $\kappa_s \to 18\alpha'$, and r_κ is continuous and increasing in κ. So for every $\mu \neq \pi_m$ with uniform marginals every exponent $c < r_{18\alpha'}/2$ is valid. Here $r_{18\alpha'}/2 \geq \max\big(8\alpha', 9\alpha'(1-9\alpha')\big)$ by Lemma 8.35: with $t = \sqrt{1-\kappa}$, $r_\kappa/\kappa = 2/(2+t(1-t)) \geq 1 - \frac{t(1-t)}{2} \geq 1 - \frac{\kappa}{2}$, as $1 - t \leq \kappa$. In the same way, Proposition 8.81(b) shows that, for $m = 3$ and two uniform marginals, every exponent below $r_{18\alpha} r_{9\alpha/(1-9\alpha)}/4 \geq 32\alpha^2$ is valid. (The bound $\mu_{ll'} \geq \frac{9\alpha}{1-9\alpha}\mu_l\mu_{l'}$ used there for Step 5 also improves the exponent $rr'/4$ of Theorem 8.61(b), for $\alpha < \frac{1}{18}$, to $r_{9\alpha} r_{9\alpha/(1-9\alpha)}/4 \geq 16\alpha^2$.) For a single $\mu \neq \pi_m$ with full support and uniform marginals, the supremum $c^*(\mu)$ of the exponents c of bounds $\mathrm{dist} \leq C\,\mathrm{err}_\mu^c$ (for every n) therefore satisfies $c^*(\mu) \geq r_{18\alpha'}/2$, and also $c^*(\mu) \leq \frac{1}{2}$: choose (a,b) with $\mu_{m-1,m}(a,b) \leq \frac{1}{9}$, and for $k \geq 1$ and $n = k$ let $f_j \equiv 0$ for $j \leq m-2$, $f_{m-1} = \mathbf{1}_E$ and $f_m = 1 + \mathbf{1}_{E'}$, where $E = \{(a,\dots,a)\}$ and $E' = \{(b,\dots,b)\}$. A failure means $\mathbf{c}_{m-1} \in E$ and $\mathbf{c}_m \in E'$, so the error is $\mu_{m-1,m}(a,b)^k \leq 9^{-k}$, while the distance from every generalized polymorphism is at least 3^{-k}: a tuple of type (B_m) has $\Pr[f_1 \neq g_1] = \frac{2}{3}$, and in a tuple of type (A_m) g_{m-1} or g_m is a constant, while $\mathbf{1}_E$ and $1 + \mathbf{1}_{E'}$ differ from every constant with probability at least 3^{-k} (for $m = 3$ read (B) and (A)). So $c^*(\mu) \to \frac{1}{2} = c^*(\pi_m)$ as $\mu \to \pi_m$ among the distributions with uniform marginals: the best exponent is continuous at π_m. Proposition 8.82 below gives the upper bound $9\alpha'$ for the class, within the factor $18\alpha'/r_{18\alpha'} \leq \frac{9}{8}$ of $r_{18\alpha'}/2$. The script `neq_large_kappa_check.py` (directory `numerics`) checks, for random distributions with uniform marginals ($m = 3, 4$) and several s, the decomposition and the bounds $\nu \geq \alpha(1-s)/(1-\sigma) \geq \frac{\alpha}{2}$, $\nu_{jl} \geq \kappa_s\nu_j\nu_l$ and $\nu_{jl} \geq \alpha'$ in exact arithmetic; Lemma 8.35 for $\kappa = \kappa_s$ on random nonnegative functions; the bound $\mu_{lm} \geq 6\alpha\,\mu_l\mu_m$ of case (b) for random μ with two uniform marginals; the error and the distance of the example for

$c^*(\mu) \le \frac{1}{2}$, by exact enumeration ($m = 3$, $k \le 3$ and $m = 4$, $k \le 2$; the distance against all tuples of types (A_m) and (B_m), and for $k = 1$ against all generalized polymorphisms, found by exhaustive search); and the numerical inequalities used for the constants and the limits.

Proposition 8.82 (The sharp constant for uniform marginals)**.** *Let $m \ge 3$, let μ be a full-support distribution on $P_{/}^{(m)} =$ with uniform marginals, and let $M = \max \mu_{jl}(a,b)$, the maximum over $j \ne l$ in $[m]$ and $a, b \in \mathbb{Z}_3$, where μ_{jl} is the law of the j-th and l-th entries of a μ-row. If constants $C, c > 0$ are such that every $\mathbf{f}$ (for every n) is within $C \operatorname{err}_\mu(\mathbf{f})^c$ of a generalized polymorphism, then $c \le \frac{9}{4}(\frac{1}{3} - M)$. In particular, for every $\alpha > 0$ with $\alpha' = 3^{m-3}\alpha \le \frac{1}{18}$, the distribution $\mu^{\mathrm{sh}}_{m,\alpha}$, defined as in Theorem 8.80(b) ($\mu^{\mathrm{sh}}_{3,\alpha} = \mu^{\mathrm{sh}}_{\alpha}$ of Proposition 8.62(b)), has full support, minimum α and uniform marginals, and admits no such bound with $c > 9\alpha'$.*

Proof. Choose $j \ne l$ and $a_0, b_0 \in \mathbb{Z}_3$ with $\mu_{jl}(a_0, b_0) = M$, and put $p = \frac{1}{3} - M$; as μ has full support, $\mu_{jl}(a_0, b) > 0$ for $b \ne b_0$, so $p > 0$. For $w \in \mathbb{Z}_3^n$ let $N_v(w)$ be the number of i with $w_i = v$. Fix τ with $0 < \tau < \min(p, \frac{1}{6})$, let $A = \{w : N_{a_0}(w) \ge (\frac{1}{3} + \tau)n\}$ and $B = \{w : N_{b_0}(w) \le (\frac{1}{3} - \tau)n\}$, and let $f_j = \mathbf{1}_A$, $f_l = 1 + \mathbf{1}_B$ and $f_k \equiv 0$ for $k \notin \{j, l\}$.

The error. A failure means $\mathbf{1}_A(\mathbf{c}_j) + 1 + \mathbf{1}_B(\mathbf{c}_l) = 0$, i.e. $\mathbf{c}_j \in A$ and $\mathbf{c}_l \in B$. Let $\mathbf{X}_i = \mathbf{1}[(\mathbf{c}_j)_i = a_0] - \mathbf{1}[(\mathbf{c}_l)_i = b_0]$; then $\sum_i \mathbf{X}_i = N_{a_0}(\mathbf{c}_j) - N_{b_0}(\mathbf{c}_l) \ge 2\tau n$ on a failure. The $\mathbf{X}_i$ are independent, with $\Pr[\mathbf{X}_i = 1] = \mu_j(a_0) - M = p$ and $\Pr[\mathbf{X}_i = -1] = \mu_l(b_0) - M = p$. With $\theta = \tau/p \in (0, 1)$, $\mathbb{E}e^{\theta \mathbf{X}_i} = 1 + 2p(\cosh\theta - 1) \le \exp(p\theta^2 \cosh\theta)$, as $\cosh\theta - 1 = \sum_{k \ge 1} \theta^{2k}/(2k)! \le \frac{\theta^2}{2}\cosh\theta$. By Markov's inequality,

$$\operatorname{err}_\mu(\mathbf{f}) \le \Pr\Big[\sum_i \mathbf{X}_i \ge 2\tau n\Big] \le e^{-2\theta\tau n}\big(\mathbb{E}e^{\theta \mathbf{X}_1}\big)^n \le e^{-\Lambda n}, \qquad \Lambda = \frac{\tau^2}{p}\Big(2 - \cosh\frac{\tau}{p}\Big) > 0.$$

The distance. The columns $\mathbf{c}_j$ and $\mathbf{c}_l$ are uniform on $\mathbb{Z}_3^n$, so $N_{a_0}(\mathbf{c}_j)$ and $N_{b_0}(\mathbf{c}_l)$ have the distribution $\mathrm{Bin}(n, \frac{1}{3})$; in particular $\Pr[A], \Pr[B] \le \frac{1}{2}$ for all large n, by the law of large numbers. Write $d(q) = d(q\|\frac{1}{3})$ as in the proof of Proposition 8.62(b), where $\Pr[\mathrm{Bin}(n, \frac{1}{3}) = k] \ge \frac{1}{n+1}e^{-nd(k/n)}$ is shown for $0 \le k \le n$. As $d(\frac{1}{3}) = d'(\frac{1}{3}) = 0$ and $d''(q) = \frac{1}{q(1-q)}$, Taylor's theorem gives $d(q) \le (q - \frac{1}{3})^2/(2\min_t t(1-t))$, the minimum over the t between $\frac{1}{3}$ and q. Let $n \ge 7$, $\tau_n = \tau + \frac{1}{n}$, $k_A = \lceil(\frac{1}{3} + \tau)n\rceil$ and $k_B = \lfloor(\frac{1}{3} - \tau)n\rfloor$. Then $|\frac{k}{n} - \frac{1}{3}| \le \tau_n$ for $k \in \{k_A, k_B\}$, $\frac{k_A}{n} \in [\frac{1}{3}, \frac{2}{3}]$, where $t(1-t) \ge \frac{2}{9}$, and $\frac{k_B}{n} \ge \frac{1}{3} - \tau_n > 0$, where $t(1-t) \ge (\frac{1}{3} - \tau_n)(\frac{2}{3} + \tau_n)$ on $[\frac{1}{3} - \tau_n, \frac{1}{3}]$, and this is at most $\frac{2}{9}$. Hence

$$\Pr[A] \ge \Pr\big[N_{a_0}(\mathbf{c}_j) = k_A\big] \ge \frac{e^{-nK_n}}{n+1}, \qquad \Pr[B] \ge \Pr\big[N_{b_0}(\mathbf{c}_l) = k_B\big] \ge \frac{e^{-nK_n}}{n+1},$$
$$K_n = \frac{\tau_n^2}{2(\frac{1}{3} - \tau_n)(\frac{2}{3} + \tau_n)}.$$

As μ has full support, the generalized polymorphisms are the tuples of types (A_m) and (B_m) (Theorem 8.70(a); types (A) and (B) of Theorem 8.55(a) for $m = 3$). A tuple $\mathbf{g}$ of type (B_m) has $\Pr[f_k \ne g_k] = \frac{2}{3}$ for $k \notin \{j, l\}$; in a tuple of type (A_m) all functions but one are constants, so g_j or g_l is a constant, and $\mathbf{1}_A$ and $1 + \mathbf{1}_B$ differ from every constant with probability at least $\min(\Pr[A], 1 - \Pr[A])$ and $\min(\Pr[B], 1 - \Pr[B])$, respectively. So for all large n, $\operatorname{dist}(\mathbf{f}, \mathbf{g}) \ge \min(\Pr[A], \Pr[B]) \ge e^{-nK_n}/(n+1)$ for every generalized polymorphism $\mathbf{g}$.

Conclusion. If $\operatorname{dist} \le C \operatorname{err}_\mu^c$ for these tuples, then $e^{-nK_n}/(n+1) \le Ce^{-c\Lambda n}$ for all large n; taking logarithms, dividing by n and letting $n \to \infty$ gives $c\Lambda \le \lim_n K_n$, i.e.

$$c \le \frac{p}{2(\frac{1}{3} - \tau)(\frac{2}{3} + \tau)(2 - \cosh\frac{\tau}{p})}.$$

Letting $\tau \to 0$ gives $c \le \frac{p}{2\cdot\frac13\cdot\frac23} = \frac94 p$.

For $\mu^{\mathrm{sh}}_{m,\alpha}$: its larger mass $\frac{1-4\cdot3^{m-2}\alpha}{2\cdot3^{m-2}}$ is at least α iff $\alpha' \le \frac{1}{18}$, and the translations used in the proofs of Proposition 8.62(b) and Theorem 8.80(b) show that its marginals are uniform; these arguments use only $\alpha' \le \frac{1}{18}$. Every pair of entries has $2\cdot3^{m-3}$ completions in $P^{(m)}_{/}$ $=$, and the completions of $(y_{m-1}, y_m) = (b, b+1)$ all have the larger mass, so $\mu_{m-1,m}(b,b+1) = 2\cdot3^{m-3}\cdot\frac{1-4\cdot3^{m-2}\alpha}{2\cdot3^{m-2}} = \frac{1-12\alpha'}{3}$. Hence $M \ge \frac{1-12\alpha'}{3}$ and $c \le \frac94\big(\frac13 - \frac{1-12\alpha'}{3}\big) = 9\alpha'$. □

For every full-support μ with uniform marginals and $\min\mu = \alpha$, $\mu_{jl}(a,b) = \frac13 - \sum_{b'\ne b}\mu_{jl}(a,b') \le \frac13 - 4\alpha'$, as each $\mu_{jl}(a,b')$ is a sum of $2\cdot3^{m-3}$ masses at least α. So in this class the bound of Proposition 8.82 is at least $9\alpha'$, with equality for $\mu^{\mathrm{sh}}_{m,\alpha}$. Together with Proposition 8.81: for every $m \ge 3$ and $0 < \alpha' < \frac{1}{18}$, the supremum of the exponents c such that every full-support μ on $P^{(m)}_{/}$ $=$ with uniform marginals and $\min\mu = \alpha$ admits a bound $\mathrm{dist} \le C_\mu\,\mathrm{err}_\mu^c$ lies in

$$\Big[\frac{r_{18\alpha'}}{2},\ 9\alpha'\Big] \subseteq \big[9\alpha'(1-9\alpha'),\ 9\alpha'\big].$$

With $t = \sqrt{1-18\alpha'}$, the ratio of the two ends is $18\alpha'/r_{18\alpha'} = 1 + \frac{t(1-t)}{2} \le \frac98$, with equality only for $\alpha' = \frac{1}{24}$, and it tends to 1 as $\alpha' \to 0$ and as $\alpha' \to \frac{1}{18}$. This replaces the upper bound $\frac92 e^2\alpha'$ of Proposition 8.62(b) and Theorem 8.80(b), proved there for $\alpha' \le \frac{1}{45}$, by $9\alpha'$ for every α'; as $9\alpha' \le \log 3/\log\frac{1}{2\alpha'}$ for $0 < \alpha' \le \frac{1}{18}$ (the function $x \mapsto 9x\ln\frac{1}{2x}$ is increasing on $(0, \frac{1}{2e})$ and equals $\ln 3$ at $\frac{1}{18}$), neither the family $\mu^{(\alpha)}$ of Proposition 8.62(a) nor its analogue $\mu^{(\alpha)}_m$ of Remark 8.71(c), whose bound is $\log 3/\log\frac{1}{2\alpha'}$, improves it. In particular the constant 9 is sharp: the supremum is $9\alpha'(1-O(\alpha'))$ as $\alpha' \to 0$. For a single $\mu \ne \pi_m$ with full support and uniform marginals, $r_{18\alpha'}/2 \le c^*(\mu) \le \frac94(\frac13 - M)$, and $\frac94(\frac13 - M) \le \frac12$ as $M \ge \frac19$. As $\mu^{\mathrm{sh}}_{m,\alpha}$ has full support and uniform marginals, the exponents of Theorems 8.61(b), 8.66 and 8.78 can be improved at most to $9\alpha'$.

The two ends of the interval are the two sides of reverse hypercontractivity for the law of (y_{m-1}, y_m) under $\mu^{\mathrm{sh}}_{m,\alpha}$. The smallest ratio $\mu_{m-1,m}(a,b)/(\mu_{m-1}(a)\mu_m(b))$ is $9\cdot2\alpha' = 18\alpha'$, attained for $b \ne a+1$, so for $\alpha' < \frac{1}{18}$ Lemma 8.35 applies with $\kappa = 18\alpha'$ and gives $\Pr[\mathbf{c}_{m-1} \in A, \mathbf{c}_m \in B] \ge (\Pr[A]\Pr[B])^{1/r_{18\alpha'}}$ for all sets A, B. For the sets of the proof (with $j = m-1$, $l = m$, $a_0 = b$, $b_0 = b+1$ and $p = 4\alpha'$), the bounds there give $\Pr[\mathbf{c}_{m-1} \in A, \mathbf{c}_m \in B] \le e^{-\Lambda n} \le \big((n+1)^2\Pr[A]\Pr[B]\big)^{\Lambda/(2K_n)}$, and $\Lambda/(2\lim_n K_n) \to \frac{1}{18\alpha'}$ as $\tau \to 0$. As $\Pr[A], \Pr[B] \le e^{-2\tau^2 n}$ by Hoeffding's inequality, the exponent $1/r_\kappa$ of Lemma 8.35 cannot be replaced by any $\gamma < 1/\kappa$ for these pair laws, and the factor $\kappa/r_\kappa \le \frac98$ is the only gap left in the interval. Proposition 8.83 removes it for $\alpha' \le \frac{1}{36}$. The script `neq_sharp_shift_check.py` (directory `numerics`) checks, in exact arithmetic, the masses, support, minimum and marginals of $\mu^{\mathrm{sh}}_{m,\alpha}$ for $m = 3, 4, 5$ and several $\alpha' \le \frac{1}{18}$, that $M = \frac{1-12\alpha'}{3}$ (attained by the last two entries), the probabilities $\Pr[\mathbf{X}_i = \pm1] = 4\alpha'$ and the smallest ratio $18\alpha'$; for random μ with uniform marginals ($m = 3, 4$), the bounds $\frac19 \le M \le \frac13 - 4\alpha'$ and $\Pr[\mathbf{X}_i = \pm1] = \frac13 - M$; the inequalities for cosh, the bound $\Pr[\mathrm{Bin}(n,\frac13) = k] \ge \frac{e^{-nd(k/n)}}{n+1}$ for $n \le 200$ and the Taylor bound for d on a grid; for $m = 3$, μ^{sh}_α (four values of α, two of τ) and $n = 30, 60, \dots, 150$, the error, summed in floating point over the four classes of rows given by $\mathbf{1}[y_2 = 0]$ and $\mathbf{1}[y_3 = 1]$, against $e^{-\Lambda n}$, and $\Pr[A]$, $\Pr[B]$, exactly, against $e^{-nK_n}/(n+1)$; for $n \le 4$ the exact class formula for the error against enumeration of all row tuples; for $m = 3$ and $3 \le n \le 7$, the distance from all tuples of types (A) and (B), by exact enumeration; and the limits and inequalities for $r_{18\alpha'}$, $\log 3/\log\frac{1}{2\alpha}$ and $\Lambda/(2K)$ on grids.

Proposition 8.83 (The sharp exponent for $\alpha' \le \frac{1}{36}$)**.** *For $0 < \kappa < 1$ and a probability space Ω consider the inequality*

$$(\mathrm{S}_\kappa) \qquad (1-\kappa)\,\mathbb{E}[\varphi\psi] + \kappa\,\mathbb{E}[\varphi]\,\mathbb{E}[\psi] \;\ge\; \|\varphi\|_\kappa\,\|\psi\|_\kappa$$

for all measurable $\varphi, \psi\colon \Omega \to [0,\infty)$ with finite means.

(a) (S_κ) holds on every probability space for every $\kappa \in (0, \frac{1}{2}]$.

(b) Let $m \ge 3$ and $0 < \alpha' < \frac{1}{18}$. Every exponent $c < \max\big(\min(9\alpha', \frac{1}{4}),\, r_{18\alpha'}/2\big)$ is valid in Proposition 8.81(a). In particular, for $0 < \alpha' \le \frac{1}{36}$ the supremum of the exponents valid for the whole class (which lies in $[r_{18\alpha'}/2, 9\alpha']$ by the paragraph after Proposition 8.82) equals $9\alpha'$.

(c) If (S_κ) holds on $\mathbb{Z}_3$ with the uniform measure for every $\kappa \in (\frac{1}{2}, 1)$, then this supremum equals $9\alpha'$ for every $0 < \alpha' < \frac{1}{18}$.

Proof. (a) Let $\beta = \kappa/(1-\kappa)$, so $0 < \beta \le 1$. If $\mathbb{E}\psi = 0$ both sides of (S_κ) vanish; as both sides are homogeneous in ψ, we may assume $\mathbb{E}\psi = 1$. Let $V = (1-\kappa)\psi + \kappa$; then $V \ge \kappa$ and $\mathbb{E}V = 1$, so $1/V$ and $V^{-\beta}$ are bounded and $V^\beta \le 1 + V$ is integrable. The left side of (S_κ) is $\mathbb{E}[\varphi V]$. By Hölder's inequality with the exponents $1/\kappa$ and $1/(1-\kappa)$, $\mathbb{E}[\varphi^\kappa] = \mathbb{E}[(\varphi V)^\kappa V^{-\kappa}] \le (\mathbb{E}[\varphi V])^\kappa (\mathbb{E}[V^{-\beta}])^{1-\kappa}$, that is, $\mathbb{E}[\varphi V] \ge \|\varphi\|_\kappa (\mathbb{E}[V^{-\beta}])^{-1/\beta}$ (the reverse Hölder inequality). So it suffices to show $\mathbb{E}[\psi^\kappa]\,(\mathbb{E}[V^{-\beta}])^{1-\kappa} \le 1$, which is $(\mathbb{E}[V^{-\beta}])^{-1/\beta} \ge \|\psi\|_\kappa$. Hölder's inequality with the same exponents gives

$$\mathbb{E}[\psi^\kappa] = \mathbb{E}\big[(\psi/V)^\kappa\, V^\kappa\big] \le \big(\mathbb{E}[\psi/V]\big)^\kappa \big(\mathbb{E}[V^\beta]\big)^{1-\kappa}.$$

Let $H = \mathbb{E}[1/V]$ and $h = H - 1$; then $h \ge 0$ by Jensen's inequality, and as $\psi = (V-\kappa)/(1-\kappa)$, $\mathbb{E}[\psi/V] = (1-\kappa H)/(1-\kappa) = 1 - \beta h$, which is nonnegative. Let V' be an independent copy of V and $Z = \log V - \log V'$. As Z and $-Z$ have the same law, $\mathbb{E}[V^\beta]\,\mathbb{E}[V^{-\beta}] = \mathbb{E}[e^{\beta Z}] = \mathbb{E}[\cosh \beta Z]$ and $H = \mathbb{E}[V']\,\mathbb{E}[1/V] = \mathbb{E}[e^{-Z}] = \mathbb{E}[\cosh Z]$. Comparing power series term by term, $\cosh \beta z - 1 = \sum_{k\ge 1} \beta^{2k} z^{2k}/(2k)! \le \beta^2(\cosh z - 1)$ for real z, as $\beta \le 1$; hence $\mathbb{E}[V^\beta]\,\mathbb{E}[V^{-\beta}] \le 1 + \beta^2 h$. By Bernoulli's inequality (concavity of $x \mapsto x^\beta$), $(1-\beta h)^\beta \le 1 - \beta^2 h$. Therefore

$$\Big(\mathbb{E}[\psi^\kappa]\,(\mathbb{E}[V^{-\beta}])^{1-\kappa}\Big)^{1/(1-\kappa)} \le (1-\beta h)^\beta\,\mathbb{E}[V^\beta]\,\mathbb{E}[V^{-\beta}] \le (1-\beta^2 h)(1+\beta^2 h) \le 1.$$

(b) The proof of Proposition 8.81 uses the exponent $r = r_{\kappa_s}$ only through Step 3, where Lemma 8.35 is applied to i.i.d. pairs with a law ν_{jl} that has uniform marginals and satisfies $\nu_{jl} \ge \kappa_s/9$; the other steps hold for every $r \in (0,1]$, and the conclusion has the exponent $r/2$ (the inequality $r < \kappa_s$ noted at the start of that proof is not used, and $r \ge 8\alpha'$ also holds for the values of r below, as $8\alpha' < \frac{1}{2}$; Step 3 itself uses, besides the two-function inequality, only $r \le 1$). We show that Step 3 also holds with $r = \kappa$ for every $\kappa \le \kappa_s$ such that (S_κ) holds on $\mathbb{Z}_3$ with the uniform measure; by (a), this applies to $\kappa = \min(\kappa_s, \frac{1}{2})$. For such κ we have $\nu_{jl} \ge \kappa/9$, so $3(\nu_{jl} - \kappa/9)/(1-\kappa)$ is a doubly stochastic 3×3 matrix, and by the Birkhoff–von Neumann theorem $\nu_{jl} = \sum_\pi \lambda_\pi \nu^\pi$, a convex combination over the permutations π of $\mathbb{Z}_3$, where ν^π is the law of $(\mathbf{x}, \pi(\mathbf{y}))$ for the simple coupling $(\mathbf{x}, \mathbf{y})$ with parameter κ: $\mathbf{x}$ is uniform, and $\mathbf{y} = \mathbf{x}$ with probability $1-\kappa$ and $\mathbf{y}$ is an independent uniform element otherwise. By (S_κ) applied to φ and $\psi \circ \pi$, $\mathbb{E}_{\nu^\pi}[\varphi(x)\psi(y)] \ge \|\varphi\|_\kappa \|\psi\|_\kappa$. This extends to products $\nu^{\pi_1} \otimes \cdots \otimes \nu^{\pi_N}$ by the usual induction: for fixed last pair (x_N, y_N) apply the inequality for $N-1$ pairs, and then the inequality for one pair to $x_N \mapsto \|\Phi(\cdot, x_N)\|_\kappa$ and $y_N \mapsto \|\Psi(\cdot, y_N)\|_\kappa$, whose κ-norms are $\|\Phi\|_\kappa$ and $\|\Psi\|_\kappa$. As $\nu_{jl}^{\otimes N}$ is the

mixture of these products with weights $\prod_i \lambda_{\pi_i}$, and the right side does not depend on $(\pi_1, \dots, \pi_N)$, Step 3 holds with $r = \kappa$. So the proof of Proposition 8.81(a) gives the exponent $\min(\kappa_s, \frac{1}{2})/2$ for every admissible s, besides $r_{\kappa_s}/2$. As $\kappa_s \to 18\alpha'$ for $s \to 0$, every $c < \max(\min(9\alpha', \frac{1}{4}), r_{18\alpha'}/2)$ is valid. For $\alpha' \le \frac{1}{36}$ this maximum is $9\alpha'$, as $r_{18\alpha'} < 18\alpha'$, and Proposition 8.82 shows that no exponent above $9\alpha'$ is valid for $\mu^{\mathrm{sh}}_{m,\alpha}$.

(c) The same argument with $\kappa = \kappa_s$ gives the exponents $\kappa_s/2$, which tend to $9\alpha'$. □

Remark 8.84. The left side of (S_κ) is $\mathbb{E}[\varphi(\mathbf{x})\psi(\mathbf{y})]$ for the simple coupling, in which $\mathbf{y} = \mathbf{x}$ with probability $1-\kappa$ and $\mathbf{y}$ is an independent copy of $\mathbf{x}$ otherwise, and the proof of (a) shows that (S_κ) is equivalent to the one-function inequality $\|(1-\kappa)\psi + \kappa\mathbb{E}\psi\|_{-\kappa/(1-\kappa)} \ge \|\psi\|_\kappa$ for all $\psi \ge 0$ (for the converse take $\varphi = ((1-\kappa)\psi + \kappa\mathbb{E}\psi)^{-1/(1-\kappa)}$, which gives equality in the reverse Hölder inequality). On finite spaces, Lemma 8.35 gives (S_κ) only with the smaller exponent r_κ in place of κ. No exponent larger than κ is possible: for $\varphi = 1 + \varepsilon u$ and $\psi = 1 - \varepsilon u$ with u bounded and $\mathbb{E}u = 0$, the left side is $1 - (1-\kappa)\varepsilon^2\mathbb{E}u^2$, while $\|\varphi\|_p\|\psi\|_p = 1 - (1-p)\varepsilon^2\mathbb{E}u^2 + O(\varepsilon^3)$; so (S_κ) is tight to second order at constants. On the uniform two-point space the analogous inequality with the exponent κ is Borell's sharp reverse hypercontractive inequality (with $p = q = \kappa$ and correlation $1-\kappa$). In the one-function form, (S_κ) is $\|T_t f\|_q \ge \|f\|_p$ for the simple semigroup $T_t = e^{-t}\,\mathrm{Id} + (1-e^{-t})\mathbb{E}$ with $p = \kappa$, $q = -\kappa/(1-\kappa)$ and $e^{-t} = 1-\kappa$, that is, $t = \frac{1}{2}\log\frac{1-q}{1-p}$, Borell's time; here q is the Hölder conjugate of p, which is why the one-function and two-function forms are equivalent. For simple semigroups on arbitrary probability spaces [22] proves $\|T_t f\|_q \ge \|f\|_p$ for $t \ge \log\frac{1-q}{1-p}$, and their sharper bound for simple operators ([22, Theorem 1.12] in the arXiv version, composed through $\|\cdot\|_0$) gives, for these pairs, $t \ge 2\log\frac{2-\kappa}{2(1-\kappa)}$, the time behind the exponent r_κ of Lemma 8.35; this still exceeds Borell's time $-\log(1-\kappa)$ for $\kappa > 0$, as $(2-\kappa)^2 > 4(1-\kappa)$. Part (a) shows that Borell's time suffices for these pairs (p, q) when $\kappa \le \frac{1}{2}$. By the paragraph after Proposition 8.82, the exponent κ in (S_κ) could not be increased for the purpose of (b).

We conjecture that (S_κ) holds on $\mathbb{Z}_3$ with the uniform measure (and, possibly, on every probability space) for every $\kappa \in (\frac{1}{2}, 1)$. For $\kappa > \frac{1}{2}$ we have $\beta > 1$, and the bound $\cosh\beta z - 1 \le \beta^2(\cosh z - 1)$ used in (a) fails for small $z \ne 0$. The script `neq_rhc_sharp_check.py` (directory `numerics`) found no violation of (S_κ) on $\mathbb{Z}_3$ in a numerical search (random starts and local optimization, for κ between 10^{-3} and 0.99), while the same search finds violations when the exponent κ is replaced by 1.02κ. It found no violation of the one-function form on spaces with two, three or four points and arbitrary weights either, and it checks the Birkhoff decomposition above for random laws ν_{jl} and the steps of the proof of (a) on random finite spaces. For $\kappa \in (\frac{1}{2}, 1)$ this is numerical evidence only; only part (c) uses (S_κ) for such κ.

References

[1] N. Alon, I. Dinur, E. Friedgut and B. Sudakov, *Graph products, Fourier analysis and spectral techniques*, Geom. Funct. Anal. 14 (2004), 913–940.

[2] Y. Alekseev and Y. Filmus, *Approximate polymorphisms of predicates*, arXiv:2506.12155v2 (2025); theorem numbers refer to v2.

[3] R. Ahlswede and P. Gács, *Spreading of sets in product spaces and hypercontraction of the Markov operator*, Ann. Probab. 4 (1976), 925–939.

[4] M. Bellare, D. Coppersmith, J. Håstad, M. Kiwi and M. Sudan, *Linearity testing in characteristic two*, IEEE Trans. Inform. Theory 42 (1996), 1781–1795.

[5] A. Bonami, *Étude des coefficients de Fourier des fonctions de $L^p(G)$*, Ann. Inst. Fourier 20 (1970), 335–402.

[6] G. Chase, Y. Filmus, D. Minzer, E. Mossel and N. Saurabh, *Approximate polymorphisms*, STOC 2022; arXiv:2106.00093.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Wiley, 2006.

[8] E. Dokow and R. Holzman, *Aggregation of binary evaluations for truth-functional agendas*, Soc. Choice Welf. 32 (2009), 221–241.

[9] Y. Filmus, *Classification aggregation: a quantitative impossibility theorem*, arXiv:2605.17136 (2026).

[10] Y. Filmus, *Triviality of promise polymorphisms*, arXiv:2607.27057 (2026).

[11] Y. Filmus, N. Lifshitz, D. Minzer and E. Mossel, *AND testing and robust judgement aggregation*, STOC 2020, 222–233; arXiv:1911.00159.

[12] Y. Filmus, N. Lifshitz, D. Minzer and E. Mossel, *AND testing*, note dated April 12, 2021, https://yuvalfilmus.cs.technion.ac.il/Papers/SimpleAND.pdf.

[13] M. Ghandehari and H. Hatami, *Fourier analysis and large independent sets in powers of complete graphs*, J. Combin. Theory Ser. B 98 (2008), 164–172.

[14] D. Greenwell and L. Lovász, *Applications of product colouring*, Acta Math. Acad. Sci. Hungar. 25 (1974), 335–340.

[15] T. E. Harris, *A lower bound for the critical probability in a certain percolation process*, Proc. Cambridge Philos. Soc. 56 (1960), 13–20.

[16] N. Keller, *On the probability of a rational outcome for generalized social welfare functions on three alternatives*, J. Combin. Theory Ser. A 117 (2010) 389–410.

[17] N. Keller, *A tight quantitative version of Arrow's impossibility theorem*, J. Eur. Math. Soc. 14 (2012) 1331–1355.

[18] G. Kindler, N. Kirshner and R. O'Donnell, *Gaussian noise sensitivity and Fourier tails*, Israel J. Math. 225 (2018), 71–109.

[19] C. F. Laywine and G. L. Mullen, *Discrete Mathematics Using Latin Squares*, Wiley, New York, 1998.

[20] B. D. McKay and I. M. Wanless, *A census of small Latin hypercubes*, SIAM J. Discrete Math. 22 (2008), 719–736.

[21] E. Mossel, R. O'Donnell, O. Regev, J. E. Steif and B. Sudakov, *Non-interactive correlation distillation, inhomogeneous Markov chains, and the reverse Bonami–Beckner inequality*, Israel J. Math. 154 (2006), 299–336; arXiv:math/0410560.

[22] E. Mossel, K. Oleszkiewicz and A. Sen, *On reverse hypercontractivity*, Geom. Funct. Anal. 23 (2013), 1062–1097; arXiv:1108.1210.

[23] E. Mossel, *A quantitative Arrow theorem*, Probab. Theory Related Fields 154 (2012), 49–88.

[24] I. Nehama, *Approximately classic judgement aggregation*, Ann. Math. Artif. Intell. 68 (2013), 91–134.

[25] R. O'Donnell, *Analysis of Boolean Functions*, Cambridge University Press, 2014.

[26] Y. Polyanskiy and Y. Wu, *Strong data-processing inequalities for channels and Bayesian networks*, in: Convexity and Concentration, IMA Vol. Math. Appl. 161, Springer, 2017, 211–249.

[27] V. N. Potapov and D. S. Krotov, *On the number of n-ary quasigroups of finite order*, Discrete Math. Appl. 21 (2011), 575–585; arXiv:0912.5453.

[28] M. Parnas, D. Ron and A. Samorodnitsky, *Testing basic Boolean formulae*, SIAM J. Discrete Math. 16 (2002), 20–46.

[29] H. S. Witsenhausen, *On sequences of pairs of dependent random variables*, SIAM J. Appl. Math. 28 (1975), 100–113.