JOURNAL OF COMBINATORIAL THEORY, Series A 61, 1–12 (1992)

How to Guess Two Letters Correctly

RON AHARONI

Department of Mathematics, Technion-Israel Institute of Technology, 32000 Haifa, Israel

AND

RON HOLZMAN

Department of Applied Mathematics and Computer Science, The Weizmann Institute of Science, 76100 Rehovot, Israel

Communicated by the Managing Editors

Received January 15, 1990

A set U of functions from $[k]$ to $[n]$ is said to be (m, n, k)-*guessing* (where m, n, k are natural numbers and $2 \leqslant m \leqslant k$) if for every function w from a subset of size m of $[k]$ into $[n]$ there exists a function in U coinciding with w in at least two places. Let $g(m, n, k)$ denote the minimal size of an (m, n, k)-guessing set. We investigate the behavior of $g(m, n, k)$, with special attention to the case $m = k$.

1. INTRODUCTION

The most natural setting for the topic of this paper is that of codings. Our basic problem is this: how many words of length k in an alphabet of size n are necessary in order to "catch" every word in at least two places? (The choice of "two" is for its being the smallest interesting case. It is difficult enough for us not to venture on the study of higher numbers. Another natural case is $k-1$ places out of k. This case has been studied quite a lot, for example, in [4, 11]). Now, words are functions from the set of places into the alphabet, and it is the terminology of "functions" that we use. We shall take the alphabet to be $[n]$.

Let m, n, k be natural numbers such that $2 \leqslant m \leqslant k$. Denote by $\mathscr{W}(m, n, k)$ the set of functions from a subset of size m of $[k]$ into $[n]$ (these correspond to "partial words" of length m). We write $\mathscr{W}(n, k)$ for $\mathscr{W}(k, n, k)$. A subset U of $\mathscr{W}(n, k)$ is called (m, n, k)-*guessing* if every function in $\mathscr{W}(m, n, k)$ coincides with some function in U in at least two places; that is, if for every $w \in \mathscr{W}(m, n, k)$ there exist $u \in U$ and distinct $i, j \in \operatorname{dom} w$

0097-3165/92 $5.00

such that $w(i)=u(i)$ and $w(j)=u(j)$. Let $g(m, n, k)$ denote the minimal size of an (m, n, k)-guessing set. We shall write $f(n, k)$ for $g(2, n, k)$ and $h(n, k)$ for $g(k, n, k)$ (thus $f(n, k)$ is the minimal number of words covering all pairs of occurences of letters in any pair of places, and $h(n, k)$ is the minimal number of words which are needed to cover every word in two places).

The function $f(n, k)$ has been studied rather extensively, but only the values of $f(2, k)$ are known precisely [5, 6] for every k. For other values of n there is a large gap between the known lower and upper bounds (see [7–9]). We shall have very little to say about $f(n, k)$. The main results will concern the function h. We shall give upper and lower bounds on $h(n, k)$ which are close for n relatively large ($n \geqslant (k-1)^2$) and coincide infinitely often. For $n<k$ it is easy to see that $h(n, k)=n$, and thus the cases remaining open are $k \leqslant n<(k-1)^2$, where we believe that the upper bounds yield the correct value. We improve on the lower bound for $n=k$. Some bounds are also given on the values of $g(m, n, k)$ for general m.

2. A Dual Formulation

A *partial n-partition of a set* $\mathbf{S}$ is a family of n pairwise disjoint (possibly empty) subsets of $\mathbf{S}$. The elements of a partial n-partition π will be denoted by $\pi_1, ..., \pi_n$. We write $\|\pi\|=\sum_{i=1}^{n}|\pi_i|$. If $\bigcup_{i=1}^{n}\pi_i=\mathbf{S}$ then π is said to be an *n-partition.*

Given a family $\Pi=(\pi^1, ..., \pi^k)$ of partial partitions (we allow $\pi^i=\pi^j$ for $i \neq j$), a collection $\theta=(\pi_{j_1}^{i_1}, ..., \pi_{j_m}^{i_m})$ of elements of some of the π^i is called an *m-transversal* for Π if all indices i_t are distinct, and θ is a partial m-partition. A k-transversal for Π is simply called a *transversal.* If Π has no m-transversal it is called *m-tangled,* and if it has no transversal it is called *tangled.*

There is a "dual" approach to the calculation of $g(m, n, k)$, based on the following observation (which is similar to one made in [7], where the problem in its "dual" form is attributed to [10]):

Lemma 2.1. *There exists an (m, n, k)-guessing set of size l if and only if there exists an m-tangled family of size k of n- partitions of a set of size l.*

Proof. Suppose that there exists an (m, n, k)-guessing set $\mathbf{S}$ of size l. Define k n-partitions $\pi^1, ..., \pi^k$ of $\mathbf{S}$ as follows: $\pi_j^i=\{s \in \mathbf{S}: s(i)=j\}$ $(1 \leqslant i \leqslant k, 1 \leqslant j \leqslant n)$. Suppose now that $(\pi_{j_1}^{i_1}, ..., \pi_{j_m}^{i_m})$ is an m-transversal for the family $(\pi^1, ..., \pi^k)$. Then, there is no $s \in \mathbf{S}$ belonging to two sets $\pi_{j_r}^{i_r}$. But this means that the partial function w defined by $w(i_r)=j_r$ $(1 \leqslant r \leqslant m)$ is not identical in two places with any $s \in \mathbf{S}$, contradicting the fact that $\mathbf{S}$ is (m, n, k)-guessing. The reverse direction is similar.

3. Upper Bounds

Our upper bounds will all depend on the function $f(n, k)$, so we start by recounting a few properties of this function. First, two obvious facts (remember that we are assuming that $k \geqslant 2$):

LEMMA 3.1. (a) $f(n, k) \geqslant n^2$.

(b) *$f(n, k)$ is non-decreasing as a function of both n and k.*

LEMMA 3.2. *$f(n, k) = n^2$ if and only if there exist $k-2$ mutually orthogonal Latin squares of order n.*

This is probably well known and appears in its "dual" form as Theorem 8.2.1 in [2]. For completeness we give the proof.

Proof. Suppose that $\mathbf{S}_3, ..., \mathbf{S}_k$ are $k-2$ mutually orthogonal Latin squares. For $1 \leqslant i, j \leqslant n$ define functions $f_{ij} : [k] \to [n]$ by: $f_{ij}(1) = i$, $f_{ij}(2) = j$, $f_{ij}(t) = \mathbf{S}_t(i, j)$ $(3 \leqslant t \leqslant k)$. The fact that the family $\{f_{ij} : 1 \leqslant i, j \leqslant n\}$ is $(2, n, k)$-guessing follows from the Latinity and orthogonality of the $\mathbf{S}_t$. Conversely, if $\{f_{ij} : 1 \leqslant i, j \leqslant n\}$ is a $(2, n, k)$-guessing family of functions from $[k]$ to $[n]$ which satisfy (as we may assume they do) $f_{ij}(1) = i$, $f_{ij}(2) = j$, then the squares $\mathbf{S}_t$ $(3 \leqslant t \leqslant k)$ defined by $\mathbf{S}_t(i, j) = f_{ij}(t)$ are Latin and orthogonal.

A simple corollary is:

COROLLARY 3.2a. *$f(n, 2) = f(n, 3) = n^2$ for all n.*

Chowla, Erdös, and Straus proved [1] that for every l there exists $n_0(l)$ such that for $n \geqslant n_0(l)$ there exist l orthogonal Latin squares of order n. This gives:

COROLLARY 3.2b. *For every k there exists $N(k)$ such that $f(n, k) = n^2$ for all $n \geqslant N(k)$.*

Since the existence of $n-1$ orthogonal Latin squares of order n is equivalent to the existence of a projective plane of order n (see, e.g., [2, Theorem 5.2.2]), it follows that:

LEMMA 3.3. *$f(n, n+1) = n^2$ if and only if there exists a projective plane of order n.*

COROLLARY 3.3a. *$f(n, k) = n^2$ whenever n is a prime power and $n \geqslant k-1$.* (This follows from Lemmas 3.1 and 3.3.)

In the following Lemmas 3.4–3.6, $n \geqslant 2$ and k will denote the maximal size of a 2-tangled family of n-partitions of $[l]$. As explained in Section 2, there is a duality between statements about k (for given n and l) and statements about $f(n, k)$. More precisely, assuming that $l \geqslant n^2$, $k = \max\{k' : f(n, k') \leqslant l\}$.

LEMMA 3.4 [5, 6]. *For* $n = 2$, $k = \binom{l-1}{\lfloor l/2 \rfloor - 1}$.

This yields:

COROLLARY 3.4a. $f(2, k) = \log_2 k + \frac{1}{2}\log_2 \log_2 k + O(1)$.

For general n, the gap between the known upper and lower bounds is quite wide.

LEMMA 3.5. (a) [8] *If* n *is a prime power then* $k \geqslant (n+1)^{\lfloor l/n^2 \rfloor}$.

(b) [9] *If* n *is a prime power and* $n^2 - n \mid l - n$ *then* $k \geqslant n^{(l-n)/(n^2-n)}$.

(c) [8] $k \geqslant (2n)^{l/(4n^2)}$ *whenever* $l \geqslant 8n^3$.

(d) [7] *For every* n *there exists* $l(n)$ *such that* $k \geqslant (\frac{9}{8}n)^{l/(9/8n)^2}$ *whenever* $l \geqslant l(n)$.

In (b) the division requirement can be relaxed at the price of some weakening of the inequality. An upper bound on k in terms of n and l (which yields a lower bound on l in terms of n and k) was found by Poljak and Tuza [9]:

LEMMA 3.6. $k \leqslant \frac{1}{2}\binom{\lfloor 2l/n \rfloor}{\lfloor l/n \rfloor}$.

Any of the lower bounds on k given in Lemma 3.5 can easily be translated into an upper bound on $f(n, k)$. We find it convenient to state explicitly one such upper bound which could be slightly improved, but has the advantage of being simple and universal (requiring no restrictions on the values of n and k):

LEMMA 3.7. $f(n, k) \leqslant 4n^2 \lceil \log_{n+2} k \rceil$.

Proof. Assume first that n is a prime power and let $f(n, k) = l$. By Lemma 2.1 there does not exist a 2-tangled family of size k of n-partitions of $[l-1]$. So, by Lemma 3.5(a), $k > (n+1)^{\lfloor (l-1)/n^2 \rfloor}$. Thus $\lfloor (l-1)/n^2 \rfloor < \log_{n+1} k$, which is equivalent to $(l-1)/n^2 < \lceil \log_{n+1} k \rceil$, or to $l \leqslant n^2 \lceil \log_{n+1} k \rceil$.

For general n, we use the fact (which is also used in the proof of Lemma 3.5(c) in [8]) that there exists a prime p such that $n < p \leqslant 2n$ (see

[3, Theorem 418]). By Lemma 3.1(b) and the first part of this proof, we have $f(n,k) \leqslant f(p,k) \leqslant p^2 \lceil \log_{p+1} k \rceil \leqslant 4n^2 \lceil \log_{n+2} k \rceil$.

An easy construction gives an upper bound on $g(m,n,k)$ in terms of $f(n,k)$:

THEOREM 3.8. (a) $g(m,n,k) \geqslant n$.

(b) *If* $m > n$ *then* $g(m,n,k) = n$.

(c) *If* $m \leqslant n$ *and* $n = a(m-1)+b$ $(0 \leqslant b < m-1)$ *then* $g(m,n,k) \leqslant (m-b-1)f(a,k) + bf(a+1,k)$.

Proof. (a) Let $\mathbf{U}$ be a family of less than n functions from $[k]$ to $[n]$. Then for every $i \in [k]$ there exists $j = j(i) \in [n]$ which does not belong to $\{u(i) : u \in \mathbf{U}\}$. Since no function in $\mathbf{U}$ guesses two values (in fact, not even one) of the function $j(i)$, $\mathbf{U}$ is not (k,n,k)-guessing and, therefore, not (m,n,k)-guessing.

(b) Take $\mathbf{U}$ to be the set of constant functions: $\mathbf{U} = \{u_j : 1 \leqslant j \leqslant n\}$, where $u_j(i) = j$ for every $1 \leqslant i \leqslant k$. Since $m > n$, every function with domain of size m into $[n]$ must have two identical values, and hence it is "guessed" by some u_j.

(c) Divide $[n]$ into $m-1$ almost equal parts $A_i : m-b-1$ of size a and b of size $a+1$. For every $1 \leqslant i \leqslant m-1$ take a $(2, |A_i|, k)$-guessing set $\mathbf{U}_i$ of size $f(|A_i|, k)$ of functions from $[k]$ into A_i. Then $\mathbf{U} = \bigcup_{i=1}^{m-1} \mathbf{U}_i$ is an (m,n,k)-guessing set, since every function w from a subset of size m of $[k]$ into $[n]$ has two values in the same A_i, which are then "guessed" by some $u \in \mathbf{U}_i$.

Let us mention three special cases which will be discussed later:

COROLLARY 3.8a. (a) $g(m,m,k) \leqslant m-2+f(2,k)$.

(b) $h(n,n) \leqslant n-2+f(2,n)$.

(c) *If* $m-1 \mid n$ *and either*

(i) $n/(m-1) \geqslant k-1$ *and* $n/(m-1)$ *is a prime power, or*

(ii) $n/(m-1) \geqslant N(k)$ *(as introduced in Corollary* 3.2b),

then $g(m,n,k) \leqslant n^2/(m-1)$.

Proof. Part (a) follows by putting $a = b = 1$ in the theorem; (b) is a special case of (a); (c) follows from the theorem and Corollaries 3.2b and 3.3a.

More generally, we can obtain upper bounds on $g(m,n,k)$ by combining the theorem and any upper bound on $f(n,k)$. An upper bound on

$g(m, n, k)$ that does not require any restrictions on the values of the parameters is obtained via Lemma 3.7, namely:

COROLLARY 3.8b. $g(m, n, k) \leqslant 4(m-1)\lceil n/(m-1)\rceil^2 \lceil \log_{\lceil n/(m-1)\rceil+2} k \rceil$.

4. LOWER BOUNDS

THEOREM 4.1. $g(m, n, k) \geqslant n^2/(m-1)$.

It is easy to see that g is non-decreasing in the variable k. Hence it suffices to show that $g(m, n, m) = h(n, m) \geqslant n^2/(m-1)$. By Lemma 2.1 this will follow from:

THEOREM 4.1'. *If* $(\pi^1, ..., \pi^m)$ *is a tangled family of partial n-partitions of a set* $\mathbf{S}$ *then* $\max\{\|\pi^i\|: 1 \leqslant i \leqslant m\} \geqslant n^2/(m-1)$.

Proof of Theorem 4.1'. By induction on m. For $m=2$ even the stronger condition $\min\{\|\pi^1\|, \|\pi^2\|\} \geqslant n^2$ holds. For if, say, $\|\pi^1\| < n^2$ then some part π^1_j of π^1 is of size less than n. Then π^1_j misses at least one part π^2_k of π^2, and then (π^1_j, π^2_k) is a transversal for (π^1, π^2).

Assume now that $m>2$ and that the theorem is true up to m. Let $d = \min\{|\pi^i_j|: 1 \leqslant i \leqslant m,\ 1 \leqslant j \leqslant n\}$, say $d = |\pi^1_1|$. Clearly, we may assume that $d<n$. For every $1<i\leqslant m$ there are no more than d parts π^i_j intersecting π^1_1. We form a partial $(n-d)$-partition θ^i of $\mathbf{S}\backslash\pi^1_1$, by choosing any $n-d$ parts in π^i disjoint from π^1_1. If the partitions $\theta^2, ..., \theta^m$ had an $(m-1)$-transversal then together with π^1_1 we would have a transversal for $(\pi^1, ..., \pi^m)$, contrary to our assumption. Hence, by the induction hypothesis, there exists $1<i\leqslant m$ such that $\sum_{j=1}^{n-d} |\theta^i_j| \geqslant (n-d)^2/(m-2)$. Since $|\pi^i_j| \geqslant d$ for all j, d parts in $\pi^i\backslash\theta^i$ contain together at least d^2 elements. Hence $\|\pi^i\| \geqslant (n-d)^2/(m-2) + d^2$. But $\min\{(n-x)^2/(m-2) + x^2\}$ is attained at $x = n/(m-1)$ and is equal to $n^2/(m-1)$. Hence $\|\pi^i\| \geqslant n^2/(m-1)$, as required.

Together with Corollary 3.8a (c) we have:

COROLLARY 4.1a. *If* $m-1 \mid n$ *and either*

(i) $n/(m-1) \geqslant k-1$ *and* $n/(m-1)$ *is a prime power, or*

(ii) $n/(m-1) \geqslant N(k)$,

then $g(m, n, k) = n^2/(m-1)$.

We recall that $h(n, 2) = f(n, 2) = n^2$ for all n. For the next two values of $h(n, \cdot)$ we have:

COROLLARY 4.1b. (i) $h(n, 3) = \lceil n^2/2 \rceil$ *for every* n.

(ii) $h(n, 4) = \lceil n^2/3 \rceil$ *whenever* $n \geqslant 21$.

Proof. By Theorems 3.8 and 4.1 we have

$$\left\lceil \frac{n^2}{k-1} \right\rceil \leqslant h(n, k) \leqslant (k-b-1) f(a, k) + b f(a+1, k), \tag{4.1}$$

where $n = a(k-1) + b$, $0 \leqslant b < k-1$. If $a \geqslant N(k)$ then $f(a, k) = a^2$ and $f(a+1, k) = (a+1)^2$, and then the right-hand side of (4.1) equals $(k-1)a^2 + 2ab + b$. The left-hand side of (4.1) equals $(k-1)a^2 + 2ab + \lceil b^2/(k-1) \rceil$. Thus the difference between the right-hand side and the left-hand side is $b - \lceil b^2/(k-1) \rceil$, which, by a simple minimum argument, is less than or equal to $(k-1)/4$. For $k \leqslant 4$ (and still under the condition $a \geqslant N(k)$) it follows that the right-hand side and the left-hand side of (4.1) are equal.

Now, for part (i) we may clearly assume that $n \geqslant 2$ and therefore $a \geqslant 1 = N(3)$, so the above argument can be applied. For part (ii) we observe that $N(4) = 7$ (a pair of orthogonal Latin squares of order n exists whenever $n \geqslant 7$), and so the condition $a \geqslant N(4)$ becomes $\lfloor n/3 \rfloor \geqslant 7$, or $n \geqslant 21$.

Remark. Part (i) of the corollary was proved (under a different terminology and with a more ad hoc proof) in [4].

For other values of m, n, k the gap between the lower bound of this section and the upper bounds of Section 3 may be quite large. As already noted, we have firmer belief in the latter. We wish to support this view with some evidence, by improving the lower bounds in certain cases.

Let us start with a better bound on the value of $g(n, n, n) = h(n, n)$. The upper bound given by Corollary 3.8a (b) is

$$h(n, n) \leqslant n - 2 + f(2, n) = n + \log_2 n + \tfrac{1}{2} \log_2 \log_2 n + O(1)$$

(the last equality is given by Corollary 3.4a). We conjecture that, in fact, $h(n, n) = n - 2 + f(2, n)$. Theorem 4.1 yields $h(n, n) \geqslant n + 2$. We improve this to:

THEOREM 4.2. $h(n, n) > n + \frac{9}{10} \log_2 n$ *for* n *large enough.*

The proof will require a few lemmas. First, an arithmetical fact. Let x be the (unique) real root of the equation $(x-1)^3 = x^2$ (then $x \approx 3.148$). For any natural numbers q, t satisfying $2 \leqslant t \leqslant q$ write $\phi(q, t) = \binom{q+t}{\lfloor (q, t)/2 \rfloor + 2}$.

LEMMA 4.3. $\max \phi(q, t)$, *taken over all integer values of* t, $2 \leqslant t \leqslant q$ (q *fixed*), *is attained at* $t = [(x-2)/(x+2) + o(1)]\, q$.

Proof. Note first that $\phi(q, q-2i-1) < \phi(q, q-2i)$. Hence, for a fixed q, it suffices to find the maximum of the function $\theta(i) = \phi(q, q-2i)$. Write

$$\zeta(i) = \frac{\theta(i+1)}{\theta(i)} = \frac{(2q-3i-2)(2q-3i-3)(2q-3i-4)}{(i+3)(2q-2i)(2q-2i-1)}.$$

It is easy to see that $\zeta(i)$ is decreasing. Hence $\max \theta(i)$ is attained at the (unique) number $i = i(q)$ for which $\zeta(i-1) > 1$ and $\zeta(i) \leqslant 1$. Then, for some real number $r = r(q)$, $i - 1 < r \leqslant i$, one has $\zeta(r) = 1$. Write $c = c(q) = r(q)/q$. Then:

$$\zeta(r) = \frac{(2q-3cq-2)(2q-3cq-3)(2q-3cq-4)}{(cq+3)(2q-2cq)(2q-2cq-1)} = 1.$$

This implies that $\lim_{q \to \infty}(2-3c)^3 - c(2-2c)^2 = 0$ (equate the numerator and the denominator of $\zeta(r)$ and divide by q^3). From this it is seen that c is bounded away from 0, and, upon dividing by c^3, $\lim_{q \to \infty}(2/c-3)^3 - (2/c-2)^2 = 0$. Writing $2/c(q) - 2 = y(q)$, this implies that $\lim_{q \to \infty} y(q) = x$. Hence $\lim_{q \to \infty} c(q) = 2/(x+2)$ and thus $i(q) = [2/(x+2) + o(1)]q$, implying the lemma.

LEMMA 4.4. *Let* $\Gamma = (\gamma^1, ..., \gamma^v)$ *be a* t*-tangled family of partial* t*-partitions of a set* S. *Then, for every* $1 \leqslant i \leqslant v$ *there exist at least two parts in* γ^i, *say* γ^i_1 *and* γ^i_2, *with the following property: the number of partial partitions* γ^a, $a \neq i$, *having at least one part contained in* γ^i_j *is at most* $t-2$ ($j = 1, 2$).

Proof. Suppose that the lemma fails for, say, $i = 1$. Then all but one of the parts in γ^1 (again, to be specific, say it is γ^1_1) contain $t-1$ parts γ^a_j ($a \neq 1$) with distinct a's. We can then select a t-transversal $(\gamma^{a_1}_{j_1}, \gamma^{a_2}_{j_2}, ..., \gamma^{a_t}_{j_t})$ for Γ as follows. Let $a_1 = j_1 = 1$. For $2 \leqslant s \leqslant t$ choose inductively a_s, j_s so that $\gamma^{a_s}_{j_s} \subseteq \gamma^1_s$ and $a_s \notin \{a_1, ..., a_{s-1}\}$. The fact that $\gamma^{a_s}_{j_s} \subseteq \gamma^1_s$, $s = 1, ..., t$, implies that the parts $\gamma^{a_s}_{j_s}$ are disjoint, which means that they form a t-transversal.

LEMMA 4.5. *If* $\Gamma = (\gamma^1, ..., \gamma^v)$ *is a* t*-tangled family of partial* t*-partitions of* $[q+t]$ *all of whose parts are of size at least* 2, *then* $v \leqslant (t-1)\binom{q+t}{\lfloor (q-t)/2 \rfloor + 2}$.

Proof. By Lemma 4.4 each γ^i has two "special" parts, each containing parts from at most $t-2$ other partial partitions. Since all parts of γ^i are of size at least 2, one of the two special parts, call it ρ_i, is of size at most $\lfloor (q+t-2(t-2))/2 \rfloor = \lfloor (q-t)/2 \rfloor + 2$.

Consider the family $R = (\rho_i : 1 \leqslant i \leqslant v)$. We claim that R has a subfamily

A of size at least $v/(t-1)$ which is an antichain with respect to inclusion. Note that, as sets, two ρ_i may be equal, but we still regard them as distinct; therefore $(R, \subseteq)$ is not a partial order. We "correct" this by looking at the partial order $(R, \leqslant)$, defined by: $\rho_i \leqslant \rho_j$ iff $\rho_i \subset \rho_j$ or $[\rho_i = \rho_j$ and $i \leqslant j]$. We let $l(\rho_i)$ denote the number of elements ρ_j with $\rho_j \leqslant \rho_i$. By the choice of the parts ρ_i, we have $1 \leqslant l(\rho_i) \leqslant t-1$. Hence l is constant on a subfamily A of R of size at least $v/(t-1)$. Clearly, such A is an antichain as required. But then A is a Sperner family in $[q+t]$, whose elements are of size at most $\lfloor (q-t)/2 \rfloor + 2$. By the LYM inequality $|A| \leqslant \binom{q+t}{\lfloor (q-t)/2 \rfloor + 2}$ and hence $v \leqslant (t-1)\binom{q+t}{\lfloor (q-t)/2 \rfloor + 2}$.

Proof of Theorem 4.2. Write $q = q(n) = h(n, n) - n$. By Lemma 2.1 there exists a tangled family $\Pi = (\pi^1, ..., \pi^n)$ of n-partitions of $[n+q]$. Let $\Pi' = (\pi^i : \phi \notin \pi^i)$. Clearly, Π has a transversal if and only if Π' has, since a transversal of Π' can be completed to a transversal of Π by adding an empty part from each $\pi^i \in \Pi \setminus \Pi'$. Hence in what follows we may assume that $\Pi' = \Pi$, i.e., no π^i has an empty part. This implies that each π^i has at least $n-q$ singletons.

Let p be maximal for which Π has a p-transversal consisting of singletons. Then $n-q \leqslant p < n$, since Π has no n-transversal and an $(n-q)$-transversal can be obtained greedily: choose a singleton $\pi^1_{j_1}$ from π^1, then a singleton $\pi^2_{j_2}$ disjoint from it from π^2, then a singleton $\pi^3_{j_3}$ disjoint from $\pi^1_{j_1} \cup \pi^2_{j_2}$, etc. At each step $i \leqslant n-q$ we can make the required choice, since π^i has $n-q$ singletons, more than the number of singletons chosen so far.

Without loss of generality assume that $\theta = (\pi^1_1, \pi^2_2, ..., \pi^p_p)$ is a p-transversal for Π, where $\pi^i_i = \{i\}$ (remember that we are assuming that the partitioned set is $[n+q]$). For $1 \leqslant i \leqslant n$ denote by N_i the union of all singletons in π^i, and write $N_i^C = [n+q] \setminus N_i$. By the maximality of p, we have $\{p+1, ..., n+q\} \subseteq N_i^C$ for all $p < i \leqslant n$. Write $J = \bigcap \{N_i : p < i \leqslant n\} \subseteq \{1, ..., p\}$. Since $|N_i^C| \leqslant 2q$ and $\{p+1, ..., n+q\} \subseteq N_i^C$ $(p < i \leqslant n)$, we have $|N_i^C \cap \{1, ..., p\}| \leqslant q - n + p < q (p < i \leqslant n)$, and hence

$$|J| > p - (n-p)\, q \geqslant n - q - q^2. \tag{4.2}$$

W.l.o.g. assume that $\pi^i_j = \{j\}$ for $p < i \leqslant n$, $j \in J$.

Notice next that $\{p+1, ..., n+q\} \subseteq N_j^C$ for every $j \in J$. For, if π^j contained a singleton $\pi^j_r = \{i\}$ for some $i > p$ then $(\theta \setminus \{\pi^j_j\}) \cup \{\pi^j_r\} \cup \{\pi^{p+1}_j\}$ would be a $(p+1)$-transversal of singletons, contradicting the maximality of p. Now, for every $j \in J$ at most p parts of π^j intersect $\{1, ..., p\}$, and hence at least $n-p$ parts are contained in $\{p+1, ..., n+q\}$, and by the above all these parts are of size at least 2. Write $t = n-p$.

For each $j \in J$ choose a set ξ^j of t parts in π^j contained in

$\{p+1, ..., n+q\}$. Let $\Xi = (\xi^j : j \in J)$. Assume, for contradiction, that Ξ has a t-transversal $(\xi^{j_1}_{r_1}, ..., \xi^{j_t}_{r_t})$. Then

$$(\theta \setminus \{\pi^{j_1}_{j_1}, ..., \pi^{j_t}_{j_t}\}) \cup \{\xi^{j_1}_{r_1}, ..., \xi^{j_t}_{r_t}\} \cup \{\pi^{p+1}_{j_1}, \pi^{p+2}_{j_2}, ..., \pi^{n}_{j_t}\}$$

is a transversal for Π, contradicting the assumption that Π is tangled. Thus Ξ is t-tangled. By Lemma 4.5 $|J| \leqslant (t-1)\binom{q+t}{\lfloor (q-t)/2 \rfloor + 2}$, and by (4.2) it follows that $n < q + q^2 + (t-1)\binom{q+t}{\lfloor (q-t)/2 \rfloor + 2}$. Using Lemma 4.3 and the fact that $t \leqslant q$, we have

$$n < q + q^2 + (q-1)\binom{[2x/(x+2) + o(1)]\, q}{[2/(x+2) + o(1)]\, q}. \tag{4.3}$$

The entropy approximation formula to the binomial coefficients says that $\binom{l}{\alpha l} = 2^{l[H(\alpha)+o(1)]}$, where $H(\alpha)$ is the entropy function $H(\alpha) = -[\alpha \log_2 \alpha + (1-\alpha)\log_2(1-\alpha)]$. Thus (4.3) yields $n < 2^{l[H(\alpha)+o(1)]}$, where $l = 2x/(x+2)\, q$ and $\alpha = 1/x$. This implies

$$q > \{2/(x+2)[x \log_2 x - (x-1)\log_2(x-1)]\}^{-1} \log_2 n + o(\log_2 n).$$

A computation shows that $\{2/(x+2)[x \log_2 x - (x-1)\log_2(x-1)]\}^{-1} \approx 0.9067$, and the theorem follows.

COROLLARY 4.2a. *If k is large enough then for all m, $2 \leqslant m \leqslant k$, $g(m, m, k) > m + \frac{9}{10}\log_2 k$.*

Proof. This follows from the theorem, by the following assertion:

ASSERTION. $h(k, k) \leqslant g(m, m, k) + k - m$ $(2 \leqslant m \leqslant k)$.

Proof of the assertion. Choose an (m, m, k)-guessing set $\mathbf{U}$ of size $g(m, m, k)$ of functions from $[k]$ into $[m]$. Then $\mathbf{U}$, together with the constant functions $w_i \equiv i$ $(m < i \leqslant k)$, is a (k, k, k)-guessing set of functions in $\mathscr{W}(k, k, k)$.

In the case $m = k$, the lower bound in the corollary coincides with the one in the theorem. At the other extreme, however, when $m = 2$, the lower bound in the corollary is not as good as previous information: by Corollary 3.4a, $g(2, 2, k) = f(2, k) = \log_2 k + \frac{1}{2}\log_2 \log_2 k + O(1)$. More generally, it is possible to improve on the lower bound when m is small relative to k. We illustrate this for the case $m = 3$.

THEOREM 4.3. *If $g(3, 3, k) = l$ then $k \leqslant \frac{2}{3}\binom{l}{\lfloor l/2 \rfloor}$.*

(Compare this with the bound $k \geqslant \binom{l-2}{\lceil l/2 \rceil - 2}$ for the maximal k with $g(3, 3, k) \leqslant l$, given by Lemma 3.4 and Corollary 3.8a (a).)

Proof. By Lemma 2.1 there exists a 3-tangled family $\Pi = (\pi^1, ..., \pi^k)$ of 3-partitions of $[l]$. We may assume that all parts are non-empty; otherwise, replace every partition with an empty part by an arbitrary partition without empty parts and observe that the family thus obtained is still 3-tangled. We shall construct from parts in Π a Sperner family of size at least $\frac{3}{2}k$ of subsets of $[l]$, which will prove the theorem.

ASSERTION A. *The following is impossible*: $\pi^j_{t_1} \subseteq \pi^i_{s_1}$, $\pi^r_{t_2} \subseteq \pi^i_{s_2}$ $(i \neq j, i \neq r, j \neq r, s_1 \neq s_2)$.

Proof. In such case $(\pi^j_{t_1}, \pi^r_{t_2}, \pi^i_{s_3})$ is a 3-transversal for Π, where s_3 is the member of $\{1, 2, 3\}$ different from s_1, s_2.

Define a digraph $D = (\Pi, E)$ on Π by: $(\pi^i, \pi^j) \in E$ $(i \neq j)$ if π^i has two parts containing two corresponding (necessarily distinct) parts from π^j.

ASSERTION B. *If π^i has out-degree 0 in D then it has two parts that contain no part π^j_t except themselves.*

Proof. Otherwise, it must have two parts, say $\pi^i_{s_1}$ and $\pi^i_{s_2}$, that contain parts $\pi^j_{t_1}$ and $\pi^r_{t_2}$, respectively. Clearly $j \neq i$, $r \neq i$. Also $j \neq r$, since if $j = r$ then $(\pi^i, \pi^j) \in E$. But now we have a configuration ruled out by Assertion A.

ASSERTION C. *If $(\pi^i, \pi^j) \in E$ then all three parts of π^i contain no part π^r_t except themselves and parts of π^j.*

Proof. Suppose, for contradiction, that π^i_1, say, contains a part from π^r, where $r \neq i$, $r \neq j$. Since two of the parts of π^i contain parts from π^j, either π^i_2 or π^i_3 contains a part from π^j. We have again a contradiction to Assertion A.

ASSERTION D. *If $(\pi^i, \pi^j) \in E$ then π^i belongs to no other edge of D, except possibly (π^j, π^i).*

Proof. By Assertion C, it is impossible that $(\pi^i, \pi^r) \in E$ with $r \neq j$. Suppose then that $(\pi^r, \pi^i) \in E$, with $r \neq i$, $r \neq j$. Since also $(\pi^i, \pi^j) \in E$, we can obtain a chain from these three partitions, say $\pi^j_1 \subseteq \pi^i_1 \subseteq \pi^r_1$. But Assertion C, applied to (π^r, π^i), tells us that π^r_1 contains no part from π^j.

By Assertion D, the weakly connected components of D must have one of the following forms:

I. An isolated vertex.

II. A directed circuit of length 2.

III. A star consisting of one or more edges, all directed towards the center.

We construct a family R of parts in Π as follows. From each isolated vertex π^i, we take two parts with the property stated in Assertion B. From each directed circuit of length 2, we take all three parts of one of the partitions. From each star we take all parts of all the partitions except the center.

By the choice of its members and Assertion C, R is a Sperner family. From each component C in D we have chosen at least $\frac{3}{2}|C|$ parts. Thus $k \leqslant \frac{2}{3}|R| \leqslant \frac{2}{3}\binom{l}{\lfloor l/2 \rfloor}$.

ACKNOWLEDGMENTS

The authors are grateful to Shay Gueron for the communication of the problem and to Noga Alon for a fruitful discussion with the second author.

Note added in proof. L. Gargano, J. Körner, and U. Vaccaro have recently found a construction which asymptotically attains the upper bound mentioned in Lemma 3.6. This construction can be used, in conjunction with Theorem 3.8(c), to obtain a small asymptotic improvement on the upper bound on $g(m, c(m-1), k)$, with c constant.

REFERENCES

1. S. CHOWLA, P. ERDÖS, AND E. G. STRAUS, On the maximal number of pairwise orthogonal Latin squares of a given order, *Can. J. Math.* **12** (1960), 204–208.
2. J. DÉNES AND A. D. KEEDWELL, "Latin Squares and Their Applications," Academic Press, New York/London, 1974.
3. G. H. HARDY AND E. M. WRIGHT, "An Introduction to the Theory of Numbers," 3rd ed., Oxford Univ. Press, Oxford, 1954.
4. J. G. KALBFLEISCH AND R. G. STANTON, A combinatorial problem in matching, *J. London Math. Soc.* **44** (1969), 60–64.
5. G. O. H. KATONA, Two applications of Sperner type theorems (for search theory and truth functions), *Period. Math. Hungar.* **3** (1973), 19–26.
6. D. J. KLEITMAN AND J. SPENCER, Families of k-independent sets, *Discrete Math.* **6** (1973), 255–262.
7. S. POLJAK, A. PULTR, AND V. RÖDL, On qualitatively independent partitions and related problems, *Discrete Appl. Math.* **6** (1983), 193–205.
8. S. POLJAK AND V. RÖDL, Orthogonal partitions and covering of graphs, *Czechoslovak Math. J.* **30** (1980), 475–485.
9. S. POLJAK AND ZS. TUZA, On the maximum number of qualitatively independent partitions, *J. Combin. Theory Ser. A* **51** (1989), 111–116.
10. A. RÉNYI, "Foundations of Probability," Wiley, New York, 1971.
11. O. TAUSSKY AND J. TODD, Covering theorems for groups, *Ann. Soc. Polon. Math.* **21** (1948), 303–305.