OBSERVATIONAL COMPLEXITY

OF CHARACTER STRINGS

by

Joel Irvin Seiferas

S.B., Massachusetts Institute of Technology
(1972)

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

January, 1972

Signature of Author ____
Department of Electrical Engineering, January 21, 1972

Certified by ____
Thesis Supervisor

Accepted by ____
Chairman, Departmental Committee on Graduate Students

OBSERVATIONAL COMPLEXITY OF CHARACTER STRINGS

BY

JOEL IRVIN SEIFERAS

Submitted to the Department of Electrical Engineering
on January 21, 1972 in partial fulfillment of
the requirements for the Degree of Master of Science

ABSTRACT

How much information about a character string is needed to recognize the string? The answer to this question depends, of course, on what kind of information is available and what is meant by "how much."

Several alternative kinds of information about strings are considered in this thesis. In each case, however, the actual pieces of information are called reports, and each report is either a 1-report, a 2-report, a 3-report, or so on. In the main investigation of the thesis, the reports on a string are the strings obtained from the string by deleting some characters. For example, the reports on the string abba fall into four categories:

1. 4-reports: abba;
2. 3-reports: abb, aba (deleting the second b), aba (deleting the first b), bba;
3. 2-reports: ab, ab, aa, bb, ba, ba;
4. 1-reports: a, b, b, a.

For the kinds of information considered in this thesis, "how much information" refers to "how large an 'n-spectrum.'" The n-spectrum of a string is just the collection of all 1-reports, 2-reports, 3-reports, ..., n-reports on the string. Some strings are recognizable from just their 1-spectra (their 1-reports), some additional ones are recognizable from their 2-spectra (their 1-reports and their 2-reports), even more are recognizable from their 3-spectra, and so on.

The observational complexity of a string is just the least n for which the string is recognizable from its n-spectrum; i.e., in the context of this thesis, the observational complexity of a string tells just how much information is needed to recognize the string. The different kinds of information considered give rise to different versions of observational complexity. Three versions are actually investigated.

THESIS SUPERVISOR: John Joseph Donovan

TITLE: Associate Professor of Electrical Engineering

ACKNOWLEDGEMENT

The author wishes to express his appreciation to Professor John J. Donovan, his thesis supervisor, for his help and his confidence in the author, and to Miss Barbara Kohl for her patience with the special characters and her excellent typing of both the rough draft and the final manuscript. The author's wife, Diane, deserves special thanks for her inspiration at the start of the work and for her encouragement and patience throughout it.

The author wishes to thank the National Science Foundation for their fellowship support. In addition, this work was supported in part by Project MAC, an M.I.T. research program sponsored by the Advanced Research Projects Agency, Department of Defense, under Office of Naval Research Contract N-00014-70-A-0362-0001.

CONTENTS

I. INTRODUCTION

Conventional computers take in their input in a sequential manner. For example, consider the input string "parse." The observations made by a computer would normally be these:

The first character is p.
The second character is a.
The third character is r.
The fourth character is s.
The fifth character is e.

In the interest of speedier input, we might consider designing a computer that somehow makes all observations of its "input slate" at once. Such a computer might make the following observations of the input string "parse":

There is one p.
There is one a.
There is one r.
There is one s.
There is one e.

These observations alone do not carry enough information, however; each of the input strings "reaps," "rapes," "pears," "pares," "spear," and "spare" gives rise to precisely the same observations. (A computer with no knowledge of the English language might even see "rpesa" as the source of the observations.)

There is no single natural way to enhance the parallel observations made above to cut down on or eliminate the ambiguities; but one possibility is to allow the inclusion, in some prescribed way, of more than one character in each observation. For example, if we allow the inclusion (à la Chapter VII) of up to two characters in each observation by our computer, then "parse" yields the following additional observations:

There is one instance of having p ultimately followed by a.
There is one instance of having p ultimately followed by r.
There is one instance of having p ultimately followed by s.
There is one instance of having p ultimately followed by e.
There is one instance of having a ultimately followed by r.
There is one instance of having a ultimately followed by s.
There is one instance of having a ultimately followed by e.
There is one instance of having r ultimately followed by s.
There is one instance of having r ultimately followed by e.
There is one instance of having s ultimately followed by e.

In this case there can be no doubt that the observed word is "parse"; for p follows no character, a follows only a single p, r follows only a single p and a single a, and so on. Taking the input "5445" as another example, we get the following observations:

There are two 5's.
There are two 4's.
There are two instances of having 5 ultimately followed by 4.
There are two instances of having 4 ultimately followed by 5.
There is one instance of having 5 ultimately followed by 5.
There is one instance of having 4 ultimately followed by 4.

Obviously "4554" is identically observed; so we still have ambiguities. If we allow the inclusion of up to three characters in each observation, however, we get four more observations from "5445":

There are two instances of having 5 ultimately followed by 4 ultimately followed by 5.
There is one instance of having 5 ultimately followed by 4 ultimately followed by 4.
There is one instance of having 4 ultimately followed by 4 ultimately followed by 5.

Only "5445" could possibly be so observed. If we consider the number of characters allowed in each observation to be the "observational power" expended by our hypothetical computer, then we can summarize as follows: The minimum necessary observational power to unambiguously perceive "parse" is 2, and the minimum necessary observational power to unambiguously perceive "5445" is 3. With respect to our particular hypothetical computer, we might naturally say that the repsective "observational complexities" of "parse" and "5445" are 2 and 3, and we might say that "5445" is "more observationally complex" than "parse."

Actually there are a number of alternative ways in which we might allow our computer to observe more than one character position at a time, and each way induces its own measure of "observational complexity." For example, we could limit observations to contiguous areas; then the "power-two" observations of "parse" would be these:

There is one p.
There is one a.
There is one r.
There is one s.
There is one e.
There is one instance of having p immediately followed by a.
There is one instance of having a immediately followed by r.
There is one instance of having r immediately followed by s.
There is one instance of having s immediately followed by e.

Clearly only "parse" gives rise to these observations, so "parse" is of observational complexity 2 with respect to this hypothetical computer, too. In this thesis we provide and use a uniform definitional framework for the study of many of these measures of observational complexity.

In Chapter II we generally define "the n-spectrum realized by a character string" in such a way that it can be thought of as the collection of observations (by some computer of the same sort as those invented above) involving at most n many character positions. We say that a string is "n-local" if there is no other string realizing the same n-spectrum. A language is n-local if any pair of strings realizing the same n-spectrum are either both in the language or both outside the language. If it exists, the least n for which a character string or language is n-local is the "observational complexity" of that character string or language. In this thesis we are concerned primarily with the observational complexity of character strings.

Studies of observational complexity are aimed primarily at describing, for each n, in precisely what ways the correspondence between strings and the n-spectra they realize fails to be one-to-one; i.e., the emphasis is on the nature of observational ambiguities of character strings. In Chapter III, for example, where we list twenty-nine basic resluts that hold for every version of observational complexity that fits into our framework, we see that all strings realize the same 0-spectrum. The twenty-three questions raised in Chapter IV are mostly specific questions about the failures of the correspondences to be one-to-one. In the investigations of Chapters V, VI we easily give complete descriptions of these failures for all n. In Chapter VII we do fairly well only for $n \leq 4$; but the results of Sections C - E and the methods of Sections F - H in that chapter may provide some valuable tools for $n > 4$.

Background definitions and conventions are given in the appendices (Chapter IX), so the work to follow begins by formalizing the concepts central to the studies just outlined.

II. FUNDAMENTAL CONCEPTS

Character Strings

An alphabet is a finite set whose members, called characters, are at least two in number.[1] Let A be an arbitrary alphabet.

A character string (or just string) on A in a function whose domain is a positive[2] integer called the length of the string and whose values are in A. (When the identity of the alphabet is clear from context, the phrase "on A" may be omitted.) If ω is a string, then $\bar{\bar{\omega}}$ denotes the length of ω (of course), and $\omega(i)$ $(0 \leq i < \bar{\bar{\omega}})$ is called the $(i + 1)^{st}$ character of ω. We say that $\omega(i)$ occupies the $(i + 1)^{st}$ position of ω. If x is some character and n is the cardinality of $\{i \mid \omega(i) = x\}$, we say that ω contains n many x's.

Usually there is no confusion if we represent a string by just writing its values next to each other in order; e.g., abcc = $\{(0,a), (1,b), (2,c), (3,c)\}$. If we write the representations of two strings ω_1 and ω_2 next to each other in order, then the string represented by the result is called the concatenation $\omega_1\omega_2$ of the strings. (Clearly, concatenation is an associative, though not commutative, operation on strings.) If B is a set of strings, then B* is the closure of B under concatenation. It being natural to identify a string of length one with its first (and only) character, we write A* for the set of strings on the alphabet A.

---

1 The only thing one could say about a string over a "singleton alphabet" is how long it is: this is just number theory rather than linguistics.
2 By ruling out the "null string," we avoid having to state trivial exceptions to numerous results; clearly, we lose nothing of significance by this action.

Each of ω_1, ω_2, and ω_3 is a *substring* of $\omega_1\omega_2\omega_3$; ω_1 is an *initial substring*, and ω_3 is a *final substring.* It is sometimes convenient to denote consecutive repetitions of a substring by using an exponent; e.g., $abcbccc = a(bc)^2c^2 = abcbc^3$. (Parentheses are used for grouping.)

We define the *reverse* $\hat{\omega}$ of a string ω to be the string of length $\bar{\bar{\omega}}$ defined by $\hat{\omega}(i) = \omega(\bar{\bar{\omega}} - 1 - i)$.

Two strings ω_1 and ω_2 (on A) of the same length are *isomorphic* if $\omega_2 = f \circ \omega_1$ for some permutation f of A. Clearly isomorphism is an equivalence relation.

Two strings ω_1 and ω_2 are *rearrangements* of each other if, for each character x, $\{i \mid \omega_1(i) = x\}$ and $\{i \mid \omega_2(i) = x\}$ have the same cardinality. Clearly rearrangement is an equivalence relation.

If $\bar{\bar{A}} = 2$, then strings on A are called *binary character strings* (or just *binary strings*). Without loss of generality, we shall always assume that $A = \{a, b\}$ whenever $\bar{\bar{A}} = 2$; in this case, we call a and b *opposite characters*, and we write $a = -b$ and $b = -a$. The *opposite* of a binary string ω is that string $-\omega$ of length $\bar{\bar{\omega}}$ defined by

$$(-\omega)(i) = -(\omega(i)).$$

Observe that distinct binary strings are isomorphic if and only if they are opposites; for there is just one permutation of $A = \{a, b\}$ that is not the identity.

A (formal) *language* (on A) is a subset of A^*. If L is a language (on A), then the *reverse* of L is that language $\hat{L}$ defined by

$$\omega \in \hat{L} \Leftrightarrow \hat{\omega} \in L,$$

and the complement of L is that language L^c defined by

$$\omega \in L^c \Leftrightarrow \omega \in A^* \,\&\, \omega \notin L.$$

We say that a language is non-trivial if neither it nor its complement is empty. Two languages L_1 and L_2 (on A) are isomorphic if there is some permutation f of A such that

$$\omega \in L_2 \Leftrightarrow f \circ \omega \in L_1.$$

A set of binary strings is called a binary language. If L is a binary language, then the opposite of L is that binary language -L defined by

$$\omega \in -L \Leftrightarrow -\omega \in L.$$

Complete Events and Reduced Events

Let A be an alphabet not containing \$. (We intend for the meta-character \$ to mean "any character from A.") A complete event (on A) is a string on $A \cup \{\$\}$. The length of a complete event p is called its extent (extent (p)). We say that p is full if $p(0) \neq \$$ & $p(\bar{\bar{p}} - 1) \neq \$$. The cardinality of $\{i \in \bar{\bar{p}} \mid p(i) \neq \$\}$ is the order of p (order (p)). The complete event p occurs in the string ω if there is some $i \in \bar{\bar{\omega}}$ such that $\omega(i + j) = p(j)$ for all $j \in \bar{\bar{p}}$, where any character (i.e., any member of A) matches \$. The correspondence given by each such i is called an occurrence of p in ω.

Example: There are three occurrences of the complete event a\$a in the string abacaaa. They are given by $i = 0$, $i = 2$, $i = 4$.

We allow for various definitions of the set $\tilde{A}$ of reduced events (or just events) on A. The general scheme involves discarding some complete events and identifying some complete events of the same order. More formally, $\tilde{A}$ is a set of subsets of A with following properties:

(i) $P \in \tilde{A} \Rightarrow P \neq \emptyset$;

(ii) $P_1 \in \tilde{A}$ & $P_2 \in \tilde{A}$ & $P_1 \neq P_2 \Rightarrow P_1 \cap P_2 = \emptyset$;

(iii) $P \in A$ & $p_1 \in P$ & $p_2 \in P \Rightarrow$ order (p_1) = order (p_2).

The order of an event P (order (P)) is just the order of its members. An event occurs in a string if any of its members does, and its occurrences in the string are just those of its members.

An event of order n is called an n-event, and an occurrence of an n-event is called an n-report. (An occurrence of any event is called a report.) For each $n \in N$, we let $\tilde{A}_n$ denote the set of events on A of order no greater than n.

If P is a set of complete events, then the reverse of P is the set $\hat{P} = \{\hat{p} \mid p \in P\}$. We say that $\tilde{A}$ is reversible if the reverse of each event is an event.

If P is a set of complete events and f is a permutation of $A \cup \{\$\}$ such that $f(\$) = \$$, then $f \circ P = \{f \circ p \mid p \in P\}$ is a permutation of P. We say that $\tilde{A}$ is character-indiscernible if each permutation of each event is an event.

For each $n \in N$, an n-spectrum on A is a function $f\colon \tilde{A}_n \to N$. We say that a string ω on A realizes the n-spectrum f if, for each $P \in \tilde{A}_n$, f(P) is the number of occurrences of P in ω. Similarly, a spectrum on A is a function $f\colon \tilde{A} \to N$; and a spectrum f is realized by ω if, for each $P \in \tilde{A}$, f(P) is the number of occurrences of P in ω. We denote the n-spectrum realized by ω by $S_n(\omega)$, and we denote the spectrum realized by ω by $S(\omega)$.

A string ω on A is n-local if there is no other string ω′ on A such that $S_n(\omega') = S_n(\omega)$. A language L on A is n-local if there are not strings ω_1 and ω_2 on A such that $\omega_1 \in L$ & $\omega_2 \notin L$ & $S_n(\omega_1) = S_n(\omega_2)$. A string or a language is local if it is n-local for some $n \in N$.

Comments on Defining Events

Our general definition of a complete event is designed so that possible definitions of $\tilde{A}$ cover such a wide range that we can accommodate many of the conceivable senses of "observational complexity" that are based on "width of vision" (i.e., "order" of observations).

As stated above, there are two general ways of defining the events. We can rule out some complete events, and we can identify some complete events of the same order.

1) Discarding Complete Events

The most natural complete events to rule out are those that are not full. Such complete events may be interesting; in general, however, we will want to forget them, for the main purpose of the \$-value is to indicate "spaces skipped" ("gaps") within an event. In the first (degenerate) sample investigation below, we shall not limit ourselves to full events; by comparing the results with those in the second (still rather simple) sample investigation, we can see that ruling out complete events that are not full is a step in the right direction.

A variation of the above idea is to discard complete events which are not at least "partially full." For example, we might rule out complete events p not having either $p(0) \neq \$$ or $p(\bar{\bar{p}} - 1) \neq \$$. If we do not mind sacrificing reversibility, we might rule out just events p having $p(0) = \$$, say.

If we do not want to have "the ability to see past gaps," we might rule out every complete event p that contains instances of the meta-character

$. Alternatively, we might, for some reason, wish to rule out complete events that contain no instances of $ (i.e., that are strings on just A).

If we want a limited "ability to see past gaps," we might discard those complete events containing too many or too great a proportion of $'s. And alternatively, again, we might rule out complete events not containing enough $'s.

Finally, if we wish to have certain specified "blind spots," that is easiest of all to arrange. For example we might rule out just two complete events or just those complete events of certain extents.

2) Identifying Complete Events

If we wish to be "blind to gaps," we can identify complete events whose representations are the same when all instances of $ are omitted. (If we do this and discard complete events that are not full, we lay the foundation for the main sample investigation of this thesis. See Chapter VII.)

If we wish to be able "to perceive gaps but not their size," we can identify complete events whose representations are the same when instances of $ that follow instances of $ are omitted. Similarly, we might wish to be able "to perceive the size of gaps" only up to some limit.

If we wish to be "blind to ordering," then we might identify complete events that are rearrangements of each other.

If we are interested only in the occurrence of certain substrings, we might identify all those complete events involving the substrings and separately identify all those complete events not involving them.

Observational Complexity

For each definition of $\tilde{A}$, we get a measure on the observational complexity of character strings and formal languages. Let F be the partial function on languages and strings given by

$$F(x) = \begin{cases} \text{the least } n \in N \text{ such that } x \text{ is } n\text{-local, if } x \text{ is local;} \\ \text{undefined, otherwise.} \end{cases}$$

We call F(x) the observational complexity of x, where x is a language or a string. We say that F is reversible if $F(\omega_1) = F(\omega_2)$ whenever $\omega_2 = \hat{\omega}_1$ and either complexity is defined. We say that F is character-indiscernible if $F(\omega_1) = F(\omega_2)$ whenever ω_1 and ω_2 are isomorphic and either complexity is defined. (By definition, then, if F is reversible (character-indiscernible) and L_1 and L_2 are reverse (isomorphic) languages, then $F(L_1) = F(L_2)$ if either L_1 or L_2 is local.)

It is helpful and interesting to define one other partial function. Let G be the partial function on N given by

$$G(\ell) = \begin{cases} \text{the least } n \in N \text{ such that every string of length no greater} \\ \qquad \text{than } \ell \text{ is } n\text{-local, if it exists;} \\ \text{undefined, otherwise.} \end{cases}$$

(As there is no string of length zero, we always have G(0) = 0.)

III. UNIVERSAL RESULTS

There are a few simple facts that hold regardless of our particular choice of events. Those listed here shall be called "Facts" and referred to freely, by number, in Chapters V - VII.

Fact 1. No string is 0-local.

Proof. Since a 0-event contains only strings that are constantly \$, any two strings of the same length must have the same 0-spectrum. □

Fact 2. The (trivial) languages $\emptyset$ and A^* are 0-local; i.e., $F(\emptyset) = F(A^*) = 0$.

Proof. Immediate; for there is no pair of strings exactly one of which is in either of these languages. Intuitively, no observation at all is needed to decide membership in $\emptyset$ or A^*. □

Fact 3. If $n' \geq n$, then $S_{n'}(\omega) \supseteq S_n(\omega)$ for every string ω.

Proof. Immediate, as $n' \geq n \Rightarrow \tilde{A}_{n'} \supseteq A_n$. □

Fact 4. For each string ω, $S(\omega) = \bigcup \{S_n(\omega) \mid n \in N\}$.

Proof. Immediate, as every event has finite order. □

Fact 5. If a string or language is n-local and $n' \geq n$, then the string or language is n'-local.

Proof. Immediate, by Fact 3 and definition. □

Fact 6. If a string or language is not n-local and $n' \geq n$, then the string or language is not n'-local.

Proof. This is just a contrapositive of Fact 5. □

Fact 7. $G(0) = 0$.

Proof. As there is no string of length zero (see footnote 2, page 10), this holds trivially by definition of G. □

Fact 8. G is non-decreasing.

Proof. Immediate, by definition of G. □

Fact 9. If every string in a language is n-local, then the language is n-local.

Proof. If every string in L is n-local and $\omega \in L$ & $S_n(\omega') = S_n(\omega)$, then $\omega' = \omega$, so that, à fortiori, $\omega' \in L$. □

Fact 10. Every language that is not n-local contains a string that is not n-local.

Proof. This is just a contrapositive of Fact 9. □

Remark. Facts 9, 10 help justify the attention we give to the observational complexity of strings rather than languages.

Fact 11. If every string in a language L is local and $\{F(\omega) \mid \omega \in L\}$ is bounded, then $F(L) \leq \max(\{F(\omega) \mid \omega \in L\})$.

Proof. Suppose $n = \max(\{F(\omega) \mid \omega \in L\})$. By Fact 5 every member of L is n-local. By Fact 9, then, L is n-local; i.e., $F(L) \leq n$. □

Remark. We can actually have

$$F(L) < \max(\{F(\omega) \mid \omega \in L\}).$$

If, for example, we had

$$\begin{aligned} & S_2(\omega_1) = S_2(\omega_2) \\ \& \quad & F(\omega_1) = F(\omega_2) = 3 \\ \& \quad & \left[S_2(\omega) = S_2(\omega_1) \Rightarrow (\omega = \omega_1 \vee \omega = \omega_2)\right], \end{aligned}$$

then we would have

$$F(\{\omega_1, \omega_2\}) \leq 2 < 3 = \max(\{F(\omega_1), F(\omega_2)\}).$$

(By Theorem F.1 of Chapter VII, this example is actually realized with ω_1 = abba, ω_2 = baab in our third sample investigation below.)

Fact 12. Any finite language that contains local strings only is itself local.

Proof. This is just a corollary of Fact 11, for $\{F(\omega) \mid \omega \in L\}$ is certainly bounded if every member of the finite language L is local. □

Fact 13. If a string of length ℓ is not ℓ-local and $n \geq \ell$, then the string is not n-local.

Proof. Immediate, for no complete event of order (or even extent) greater than ℓ can occur in a string of length ℓ. □

Fact 14. A string of length ℓ is local iff it is ℓ-local.

Proof. By definition, a string that is ℓ-local is local.

Conversely, if a string of length ℓ is not ℓ-local but is n-local for some n, then either $n < \ell$ or $n > \ell$. The former is ruled out by Fact 6, and the latter is ruled out by Fact 13, however. Therefore, a string of length ℓ that is not ℓ-local is not local. □

Fact 15. If there is some string $\omega' \neq \omega$ such that $S(\omega') = S(\omega)$, then ω is not local.

Proof. If $\omega' \neq \omega$ & $S(\omega') = S(\omega)$, then $S_\ell(\omega') = S_\ell(\omega)$, where $\ell = \bar{\bar{\omega}}$, by Fact 4. By Fact 14, then, ω is not local. □

Remark. We cannot assert the converse of Fact 15; i.e., we might have ω failing to be local without there being a single string $\omega' \neq \omega$ such that $S(\omega') = S(\omega)$. At each level n, we might have a different string realizing the same n-spectrum as ω.

Fact 16. $G(\ell)$ is defined iff every string of length no greater than ℓ is local.

Proof. If $G(\ell)$ is defined, then every string of length no greater than ℓ is local, by definition.

Conversely, there are only finitely many strings (on a given alphabet) of length no greater than ℓ; so, if each of these strings is local, Fact 5 guarantees that $G(\ell)$ is defined. □

Fact 17. The domain of G is some initial segment of N.

Proof. If every string is local, then the domain of G is N, by Fact 16.

If there is some string that is not local, then let ℓ be the length of the shortest such string. By Fact 16, the domain of G is ℓ. □

Fact 18. The range of G is a subset of $\mathrm{range}(F) \cup \{0\}$.

Proof. If G is defined at ℓ, then

$$G(\ell) = \max\ (\{F(\omega) \mid \bar{\bar{\omega}} \leq \ell\}),$$ by

definition and Fact 5. □

Fact 19. Unions, intersections, and complements of n-local languages are n-local.

Proof. Assume $S_n(\omega_1) = S_n(\omega_2)$, and let $\mathcal{L}$ be a set of languages.

Suppose $\omega_1 \in \bigcup\mathcal{L}$ & $\omega_2 \notin \bigcup\mathcal{L}$. Take $L \in \mathcal{L}$ such that $\omega_1 \in L$. As $\omega_2 \notin \bigcup\mathcal{L}$, $\omega_2 \notin L$; hence, L is not n-local. Therefore, $\bigcup\mathcal{L}$ is n-local if every language in $\mathcal{L}$ is n-local.

Suppose $\omega_1 \in \bigcap\mathcal{L}$ & $\omega_2 \notin \bigcap\mathcal{L}$. Take $L \in \mathcal{L}$ such that $\omega_2 \notin L$. As $\omega_1 \in \bigcap\mathcal{L}$, $\omega_1 \in L$; hence, L is not n-local. Therefore, $\bigcap\mathcal{L}$ is n-local if every language in $\mathcal{L}$ is n-local.

Suppose $\omega_1 \in L^c$ & $\omega_2 \notin L^c$. Then $\omega_2 \in L$ & $\omega_1 \notin L$, so that L is not local. Therefore, L^c is n-local if L is n-local. □

Fact 20. If L is a local language, then $F(L) = F(L^c)$.

Proof. As $(L^c)^c = L$, Fact 19 guarantees that L, L^c are n-local for precisely the same values of n. □

Fact 21. $\{P \in \widetilde{A} \mid (S(\omega))\ (P) \neq 0\}$ is finite for each ω.

Proof. Let $\ell = \bar{\bar{\omega}}$. No complete event of extent greater than ℓ can occur in ω. Since there are only finitely many complete events of extent no greater than ℓ, then, there are only finitely many events that can occur in ω. □

Fact 22. $\{P \in \widetilde{A} \mid (S_n(\omega))\ (P) \neq 0\}$ is finite for each ω and each n.

Proof. This is a direct corollary of Fact 21. □

Fact 23. If $\widetilde{A}$ is reversible, then strings realize the same n-spectrum or spectrum iff their reverses do.

Proof. Immediate, as reverse events have the same order and $S(\hat{\omega})$ is related to $S(\omega)$ in the following fixed way:

$$(S(\hat{\omega}))\ (P) = (S(\omega))\ (\hat{P}).\ \square$$

Remark. The converse could be make to fail; just arrange for $S(\hat{\omega})$ to bear some other fixed relationship to $S(\omega)$.

Fact 24. If (for every n) strings realize the same n-spectrum whenever their reverses do, then F is reversible.

Proof. Suppose $F(\omega) = n$ & $F(\hat{\omega}) \neq n$. Then either $\hat{\omega}$ is not n-local or it is (n - 1)-local. It will suffice to show that just one of these must fail; as $\omega = \hat{\omega}$, the other must then fail by symmetry. So assume $\hat{\omega}$ is not n-local, and let $\omega' \neq \hat{\omega}$ be such that $S_n(\omega') = S_n(\hat{\omega})$. Then $S_n(\hat{\omega}') = S_n(\hat{\omega}) = S_n(\omega)$, so that ω is not n-local, a contradiction. □

Remark. Again, F might be reversible in a "messier" way, so that strings could realize different n-spectra even when their reverses realize the same n-spectrum.

Fact 25. If $\tilde{A}$ is reversible, then so is F.

Proof. As $\hat{\omega} = \omega$, this is a direct consequence of Facts 23, 24. □

Fact 26. If $\tilde{A}$ is character-indiscernible and f is a permutation of A, then $f \circ \omega_1$ and $f \circ \omega_2$ realize the same n-spectrum or spectrum if ω_1 and ω_2 do.

Proof. Immediate, as isomorphic events have the same order and $S(f \circ \omega)$ is related to $S(\omega)$ in the following fixed way:

$$(S(f \circ \omega))\ (P) = (S(\omega))\ (f^{-1} \circ P).\ \square$$

Remark. As in Fact 23, the converse might fail.

Fact 27. If $\widetilde{\{a, b\}}$ is character-indiscernible, then strings realize the same n-spectrum or spectrum if their opposites do.

Proof. This is an immediate corollary of Fact 26. □

Fact 28. If for each permutation f of A, $f \circ \omega_1$ and $f \circ \omega_2$ realize the same n-spectrum whenever ω_1 and ω_2 do, then F is character-indiscernible.

Proof. Suppose $F(\omega) = n$ & $F(f \circ \omega) \neq n$. Then either $f \circ \omega$ is not n-local, or it is (n - 1)-local. It will suffice to show that just one of these must fail; as the inverse of a permutation is a permutation, the other must then fail by symmetry. So assume $f \circ \omega$ is not n-local, and let $\omega' \neq f \circ \omega$ be such that $S_n(\omega') = S_n(f \circ \omega)$. Then $S_n(f^{-1} \circ \omega') = S_n(f^{-1} \circ f \circ \omega) = S_n(\omega)$] As $\omega' \neq f \circ \omega \Rightarrow f^{-1} \circ \omega' \neq f^{-1} \circ f \circ \omega$, this is a contradiction. □

Remark. As in Fact 24, the converse might fail.

Fact 29. If $\tilde{A}$ is character-indiscernible, then so is F.

Proof. As the inverse of a permutation is a permutation, this is a direct consequence of Facts 26, 28. □

IV. GENERAL QUESTIONS

In an investigation of a particular measure of observational complexity (induced by a particular definition of $\tilde{A}$), there are some questions that are always of interest and some that relate to the particular measure. Here are some of the more universal questions that might be raised:

1. Do we have $S_1(\omega_1) = S_1(\omega_2) \Rightarrow \bar{\bar{\omega}}_1 = \bar{\bar{\omega}}_2$?
(The answer can be no. If, for example, we rule out the complete events that are not full and the complete events involving the character a, then we have $S_1(ba) = S_1(b)$.)

2. Do we have $S_1(\omega_1) = S_1(\omega_2) \Leftrightarrow \omega_1$ is a rearrangement of ω_2?
($\Rightarrow$ can fail even if question 1 is answered affirmatively. If, for example, we rule out the complete events that are not full and identify the complete events $\{(0, b)\}$, $\{(0, a)\}$, then we have $S_1(ba) = S_1(bb)$. $\Leftarrow$ can fail, too; in the degenerate sample investigation of Chapter V, for example, every string is 1-local.)

3. If $\ell = \bar{\bar{\omega}}_2$, do we have $S_\ell(\omega_1) = S_\ell(\omega_2) \Leftrightarrow \omega_1 = \omega_2$?
(The answer can be no; above, for example, we had $S_1(ba) = S_1(b)$.)

4. Do we have $S(\omega_1) = S(\omega_2) \Leftrightarrow \omega_1 = \omega_2$?
(The answer can be no. If, for example, we rule out complete events that are not full and complete events of order greater than one and we identify the complete events $\{(0, b)\}$ and $\{(0, a)\}$, then we have $S(ba) = S(bb)$.)

5. Is there some "nice" condition equivalent to $S_n(\omega_1) = S_n(\omega_2)$ or to

$S(\omega_1) = S(\omega_2)$?

6. Is there some "nice" condition equivalent to ω being n-local or local?

7. Is F reversible? Is $\tilde{A}$?

8. Is F character-indiscernible? Is $\tilde{A}$?

9. Is there an easy way to compute F?

10. Is there an easy way to compute G?

11. Is F total on strings?

(I.e., is every string local? We can, in fact, have <u>no</u> string local. If, for example, we identify all complete events of each order, then $S(\omega_1) = S(\omega_2)$ whenever $\bar{\bar{\omega}}_1 = \bar{\bar{\omega}}_2$.)

12. Is F total on languages?

13. Is G total?

(The answer can be no. Above, for example, we have $S(\omega_1) = S(\omega_2)$ whenever $\bar{\bar{\omega}}_1 = \bar{\bar{\omega}}_2$; hence G is not defined on any positive integer.)

14. What is the range of F? Does it have holes? Is it bounded?

(These are probably the most significant questions we can ask; for the most interesting cases will be those in which there are strings and languages of every complexity, or at least of arbitrarily great complexity.)

15. What is the range of G? Does it have holes? Is it bounded? Is it different from the range of F?

16. Does every language that is not local contain a string that is not

local?

(This might fail, for example, for a language containing only local strings, but local strings or arbitrarily great observational complexity.)

17. How many strings (perhaps with some constraints on their composition) can realize the same n-spectrum or spectrum? How long must they be? Is there any easy way to find them?

(Above, in question 1, we had a situation where $S_1(b) = S_1(ba) = S_1(baa) = S_1(baaa) = \ldots.$)

18. Which n-spectra and spectra are actually realized?

19. How does the observational complexity of a concatenation of strings compare with the individual observational complexities of the strings?

20. How do the observational complexities of unions and intersections of languages compare with the observational complexities of the original languages?

21. How does the observational complexity of L* compare with that of the language L?

22. Can you decide from a recognition algorithm in a certain form what the observational complexity of a recursive language is? Can you decide from a generation algorithm in a certain form what the observational complexity of a recursively enumerable language is? Do any familiar types of automata correspond to definite parts of the hierarchy given by F?

23. Are "languages of interest" spread out in observational complexity, or are they all simple or all hard? Are there any surprise classifications?

V. A DEGENERATE SAMPLE INVESTIGATION

For a start, let us investigate the measure of observational complexity obtained by taking the complete events as our (reduced) events. More formally, let

$$\tilde{A} = \{\{p\} \mid p \text{ is a complete event on } A\}$$

The following proposition illustrates the type of "degenerateness" that occurs when we do not rule out complete events that are not full; we get information from the 0-spectrum of a string!

<u>Proposition 1</u>. For any string ω,

$$\bar{\bar{\omega}} = \max\ (\{\text{extent } (p) \mid (S_0(\omega))\ (\{p\}) \neq 0\}).$$

<u>Proof</u>. Immediate.□

The following proposition settles just about everything.

<u>Proposition 2</u>. Every string is 1-local.

<u>Proof</u>. Suppose not. Say $S_1(\omega_1) = S_1(\omega_2)$, where $\omega_1 \neq \omega_2$. By Fact 3, $S_0(\omega_1) = S_0(\omega_2)$; hence $\bar{\bar{\omega}}_1 = \bar{\bar{\omega}}_2$, by Proposition 1. Let $\ell = \bar{\bar{\omega}}_1 = \bar{\bar{\omega}}_2$.

As $\omega_1 \neq \omega_2$, there is some i_0 for which $\omega_1(i_0) = \omega_2(i_0)$. Define $p : \ell \rightarrow A \cup \{\$\}$ by

$$p(i) = \begin{cases} \omega_1(i_0), & \text{if } i = i_0; \\ \$, & \text{otherwise.} \end{cases}$$

Then $1 = (S_1(\omega_1))\ (p) = (S_1(\omega_2))\ (p) = 0$, a contradiction. □

It follows, for example, that F is identically 1, by Fact 1 and Proposition 2; i.e., we really have no measure at all on the observational complexity of character strings or languages.

VI. AN EASY SAMPLE INVESTIGATION

Now let us investigate the measure of observational complexity obtained by just taking the full complete events as our (reduced) events. More formally, let

$$\tilde{A} = \{\{p\} \mid p \text{ is a full complete event on } A\}.$$

First we consider 0-events.

Proposition 1. There are no 0-events.

Proof. As a complete event is a non-empty function, no full complete event can possibly have order zero. □

Corollary 2. No string or non-trivial language is 0-local.

Proof. As there are no 0-events, $S_0(\omega)$ is the empty function for every string ω. □

Next we consider 1-events.

Proposition 3. Every 1-event is a singleton whose member has extent one.

Proof. Immediate, by the definition of "full." □

Corollary 4. The number of occurrences of the 1-event $\{p\}$ in a string ω is just the cardinality of $\{i \mid \omega(i) = p(0)\}$. (I.e., $(S_1(\omega))(\{p\})$ is just the count of instances in ω of some character (namely, $p(0)$) from A.)

Proof. Immediate, by Proposition 3 and the definition of "occurrences." □

Corollary 5. For any string ω,

$$\bar{\bar{\omega}} = \sum_{P \in \tilde{A}_1} (S_1(\omega))(P).$$

Proof. Immediate from Proposition 3 and Corollary 4. □

Corollary 6. Strings realize the same 1-spectrum iff they are rearrangements of each other.

Proof. Let ω_1 and ω_2 be rearrangements of each other. By Proposition 1, there are no 0-events; and, by Proposition 3 and Corollary 4, $S_1(\omega_1)$ and $S_1(\omega_2)$ agree on 1-events. Therefore, $S_1(\omega_1) = S_1(\omega_2)$.

Conversely, suppose $S_1(\omega_1) = S_1(\omega_2)$. By Corollary 4, then, ω_1 and ω_2 are rearrangements of each other. □

Corollary 7. We can find arbitrarily many strings realizing the same 1-spectrum.

Proof. To find ℓ many strings realizing the same 1-spectrum, just find a string with at least ℓ many distinct rearrangements. By Corollary 6, these rearrangements all realize the same 1-spectrum. □

Example. The string bbbaaa has $\binom{6}{3} = \frac{6!}{3!\,3!} = 20$ distinct rearrangements, giving twenty strings realizing the 1-spectrum $\{(\{\{(0, a)\}\},3), (\{\{(0, b)\}\},3)\}$.

Corollary 8. There are strings that are not 1-local.

Proof. This is just a weakening of Corollary 7. □

Corollary 9. There are languages L such that $F(L) < F(\omega)$ for each $\omega \in L$.

Proof. For example, let L be the set of all rearrangements of the string ba; i.e. $L = \{ba, ab\}$. By Corollary 6,

$$S_1(\omega) = S_1(ba) \Leftrightarrow \omega \in L;$$

therefore, $F(L) \leq 1$ even though $F(ba) > 1$ & $F(ab) > 1$. □

Corollary 10. $F(\omega) = 1$ iff ω is a constant string.

Proof. By Corollary 2, no string is 0-local; so, by Corollary 6, $F(\omega) = 1$

iff there is no rearrangement of ω other than itself. This is clearly the case iff ω is a constant string. □

Corollary 11. Let L be a non-trivial language. Then F(L) = 1 iff there is no string in L, some of whose rearrangements are not in L.

Proof. By Corollary 2, L is not 0-local; so the corollary follows from Corollary 6. □

Example. If $L = \{\omega \mid \bar{\bar{\omega}} = 17\}$, then F(L) = 1. Similarly for $\{\omega \mid \omega$ contains exactly nine a's$\}$.

Corollary 12. $G(\ell) = 1$ iff $\ell = 1$.

Proof. There are non-constant strings of any length $\ell \geq 2$; so, by Corollary 10, $G(\ell) > 1$ if $\ell \geq 2$.

If $\ell = 0$, then $G(\ell) = 0$ by Fact 7.

Every string of length one is a constant string; so G(1) = 1, by Corollary 10. □

It turns out that we can complete this investigation by looking at 2-events.

Theorem 13. Every string is 2-local.

Proof. Let ω_1 and ω_2 be strings realizing the same 2-spectrum. Let $\ell = \bar{\bar{\omega}}_1$

$$= \sum_{P \in \tilde{A}_1} (S_1(\omega_1))\ (P) \quad \text{(by Corollary 5)}$$

$$= \sum_{P \in \tilde{A}_1} (S_2(\omega_1))\ (P)$$

$$= \sum_{P \in \tilde{A}_1} (S_2(\omega_2))\ (P)$$

$$= \sum_{P \in \tilde{A}_1} (S_1(\omega_2))\ (P)$$

$$= \bar{\bar{\omega}}_2 \qquad \text{(by Corollary 5).}$$

Now suppose $\omega_1 \neq \omega_2$. Let i_0 be the least position i such that either $\omega_1(i) \neq \omega_2(i)$ or $\omega_1(\ell - 1 - i) \neq \omega_2(\ell - 1 - i)$. Let $\{p_i\}$ and $\{q_i\}$ be the 2-events occuring at position i $(0 \le i \le i_0)$ of ω_1 and ω_2, respectively, with extent(p_i) = extent$(q_i) = \ell - i_0$. Clearly, there is no other occurrence in ω_1 or ω_2 of a 2-event $\{p\}$ with extent(p) $= \ell - i_0$.

As $\omega_1(i) = \omega_2(i)$ for $i = 0, 1, \ldots, i_0 - 1$, we must have $p_i(0) = q_i(0)$ for $i = 0, 1, \ldots, i_0 - 1$. As $S_2(\omega_1) = S_2(\omega_2)$, we can conclude that $p_{i_0}(0) = q_{i_0}(0)$, too. Then $\omega_1(i_0) = p_{i_0}(0) = q_{i_0}(0) = \omega_2(i_0)$. By considering the values $p_i(\ell - 1 - i_0)$ and $q_i(\ell - 1 - i_0)$ in a similar manner, we also get $\omega_1(\ell - 1 - i_0) = \omega_2(\ell - 1 - i_0)$. The two together contradict our choice of i_0, so we must have $\omega_1 = \omega_2$. □

Corollary 14. For each $n \ge 2$, every string and every language is n-local.

Proof. Let $n \ge 2$, and let ω and L be a string and a language, respectively. By Theorem 13, ω is 2-local; so, by Fact 5, ω is n-local. Similarly, every string in L is n-local; so, by Fact 9, L is n-local. □

Corollary 15. $F(\omega) = 2$ iff ω is not a constant string.

Proof. If $F(\omega) = 2$, then ω is not a constant string, by Corollary 10.

Conversely, suppose ω is not a constant string. By Theorem 13, $F(\omega) \le 2$; and, by Corollaries 2 and 10, $F(\omega) \neq 0$ & $F(\omega) \neq 1$. Therefore, $F(\omega) = 2$. □

Corollary 16. For each $\ell > 1$, $G(\ell) = 2$.

Proof. By Fact 8 and Corollary 12, $G(\ell) \ge 2$ if $\ell > 1$. By Theorem 13, then, $G(\ell) = 2$ if $\ell > 1$. □

Corollary 17. F is total on strings and has $\{1,2\}$ for its range.

Proof. Immediate by 10, 15. □

Corollary 18. F is total on languages and has $3 = \{0, 1, 2\}$ for its range.

Proof. Immediate by 11, 14, and Fact 2; for $F(\{\omega\}) = F(\omega)$. □

Corollary 19. G is total and has $3 = \{0, 1, 2\}$ for its range.

Proof. Immediate, by 12, 16, and Fact 7. □

It should be clear by now that we can answer any question about this particular measure of observational complexity; hence the investigation is completed. This particular measure is quite useless and uninteresting, of course, as it breaks strings and languages down into just two simple classes. The general approach however is instructive and can help to guide us through more difficult investigations, such as the one that follows.

VII. A SIGNIFICANT SAMPLE INVESTIGATION

A. Foundation

For this investigation, we again discard every complete event that is not full; in addition, we identify complete events whenever deleting instances of \$ leaves them alike. (By this latter action, we deny ourselves the ability "to perceive gaps" in our reports on strings.) To be more precise, we identify full complete events p_1 and p_2 iff both of the following hold:

(i) For each $j_1 \in N$, there is some $j_2 \in N$ such that, for each character x, the sets $\{i < j_1 \mid p_1(i) = x\}$ and $\{i < j_2 \mid p_2(i) = x\}$ have the same cardinality. (As $\$ \notin A$, we do not count \$ as a character.)

(ii) For each $j_2 \in N$, there is some $j_1 \in N$ such that, for each character x, the sets $\{i < j_1 \mid p_1(i) = x\}$ and $\{i < j_2 \mid p_2(i) = x\}$ have the same cardinality.

Clearly, every n-event contains a unique representative that is a string (of length n) on just A; and every string on A represents a unique event.

B. Short Events

To get started, we quickly investigate events of orders zero and one. We can hardly expect these trivial events to be truly representative, but an understanding of them is a natural first step.

A complete understanding of 0-events comes easily.

Proposition B.1. There are no 0-events.

Proof. As a complete event is a non-empty function, no full complete event can possibly have order zero. □

Corollary B.2. No string or non-trivial language is 0-local.

Proof. As $\tilde{A}_0 = \emptyset$, $S_0(\omega)$ is the empty function for every string ω. □

The 1-events, too, are like those in the preceeding investigation (Chapter VI).

Proposition B.3. Every 1-event is a singleton whose member has extent one.

Proof. Any full complete event containing \$'s must have order at least two. □

Just as in the preceeding investigation, this proposition yields the following corollaries:

Corollary B.4. The number of occurrences of the 1-event $\{p\}$ in a string ω is just the cardinality of $\{i \mid \omega(i) = p\ (0)\}$. (I.e., $(S_1(\omega))\ (\{p\})$ is just the count of instances in ω of the character $p(0)$ from A.)

Corollary B.5. For any string ω,

$$\bar{\bar{\omega}} = \sum_{P \in \tilde{A}_1} (S_1(\omega))\ (P).$$

Corollary B.6. Strings realize the same 1-spectrum iff they are rearrangements of each other.

Corollary B.7. We can find arbitrarily many strings realizing the same 1-spectrum.

Corollary B.8. There are strings that are not 1-local.

Corollary B.9. There are languages L such that $F(L) < F(\omega)$ for each $\omega \in L$.

Corollary B.10. $F(\omega) = 1$ iff ω is a constant string.

Corollary B.11. Let L be a non-trivial language. Then $F(L) = 1$ iff there is no string in L, some of whose rearrangements are not in L.

Corollary B.12. $G(\ell) = 1$ iff $\ell = 1$.

It is easy to verify one more corollary:

Corollary B.13. There are strings of observational complexity 2.

Proof. By B.6, ab is not 1-local, but the only other string realizing $S_1(ab)$ is ba. As $(S_2(ab))([ab]) = 1 \neq 0 = (S_2(ba))([ab])$, $S_2(ab) \neq S_2(ba)$. By Fact 3, $S_1(ab) \subseteq S_2(ab)$; so ab is 2-local. Therefore, $F(ab) = 2$. □

C. Simplifications

In various ways, we can simplify our investigation. We just look for symmetries and relationships that allow us to see that certain questions will have the same answer and that certain questions can be reduced to others in some sense. Four important dimensions of simplification are reported here. They are reversibility, character-indiscernibility, "binary sufficiency," and "cancellation." First, however, we prove an easier proposition that also fits in here.

Proposition C.1. Let ω be a string of length ℓ. If there is a string $\omega' \neq \omega$ realizing the same n-spectrum (or spectrum) as ω, then there is a string $\omega'' \neq \omega$ of length ℓ that does so.

Proof. If n = 0, then we have seen that every string realizes the same n-spectrum (namely, the empty function).

If $n > 0$, then only strings of length ℓ can realize the same n-spectrum as ω, by B.6 and Fact 3. Similarly, only strings of length ℓ can realize the same spectrum as ω. □

Corollary C.2. If ω is not local, then there is some string $\omega' \neq \omega$ of length $\ell = \bar{\bar{\omega}}$ such that $S(\omega') = S(\omega)$.

Proof. Assume ω is not local. For each $n \in N$, let $f(n) \neq \omega$ be a string of length ℓ realizing $S_n(\omega)$. There are only finitely many strings (on A) of length ℓ, so some string must occur infinitely often as a value of f. Let ω' be such a string. By Facts 3 and 4, we must have $S(\omega') = S(\omega)$. □

This gives us the converse of Fact 15, so that we have C.3:

Corollary C.3. A string fails to be local iff there is some other string realizing the same spectrum.

<u>Proposition C.4</u>. If $P = [\omega]$, then $\hat{P} = [\hat{\omega}]$.

<u>Proof</u>. Clearly, we get the same string if we drop all instances of \$ and then reverse a string as if we reverse the string and then drop all instances of \$; Therefore, $P \in [\omega] \Leftrightarrow \hat{P} \in [\hat{\omega}]$.□

<u>Corollary C.5 (Reversibility)</u>. $\tilde{A}$ is reversible.

<u>Proof</u>. By C.4, if $P = [\omega]$, then $\hat{P}$ is just the event $[\hat{\omega}]$. □

Facts 23, 25 immediately give us two corollaries to reversibility:

<u>Corollary C.6</u>. Strings realize the same n-spectrum or spectrum iff their reverses do.

<u>Corollary C.7</u>. Reverse strings have the same observational complexity; i.e., F is reversible.

Reversibility enables us to take practically any result we get about particular strings and to transform it into a similar result about the strings' reverses. For most purposes, it will make no difference "from which end" we observe strings.

<u>Proposition C.8</u>. Let f be a permutation of $A \cup \{\$\}$ such that $f(\$) = \$$. If $P = [\omega]$, then $f \circ P = [f \circ \omega]$.

<u>Proof</u>. As $f(\$) = \$$, the operation "drop all instances of \$" commutes with f on complete events: therefore, $P \in [\omega] \Leftrightarrow f \circ p \in [f \circ \omega]$. □

<u>Corollary C.9. (Character-indiscernibility)</u>. $\tilde{A}$ is character-indiscernible.

<u>Proof</u>. Let f be a permutation of $A \cup \{\$\}$ such that $f(\$) = \$$. By C.8, if $P = [\omega]$, then $f \circ P$ is just the event $[f \circ \omega]$. □

Facts 26, 27, 29 immediately give us four corollaries to character-indiscernibility:

Corollary C.10. If f is a permutation of A, then $f \circ \omega_1$ and $f \circ \omega_2$ realize the same n-spectrum or spectrum if ω_1 and ω_2 do.

Corollary C.11. Binary strings realize the same n-spectrum or spectrum if their opposites do.

Corollary C.12. Isomorphic strings have the same observational complexity; i.e., F is character-indiscernible.

Corollary C.13. Opposite binary strings have the same observational complexity.

By now it should be clear that character-indiscernibility enables us to take practically any result we get about particular strings and to transform it into a similar result about ("simultaneously") isomorphic strings. We lose no generality, then, if we "rename" the characters as we please or if we make an arbitrary choice of which of several isomorphic strings (or languages) we study.

Our next result enables us to limit a great deal of our investigation to binary strings. In particular, it shows us that G does not depend on the size of the alphabet.

Theorem C.14 (Binary Sufficiency). If there is a string of length ℓ that fails to be n-local or local, then there is a binary string of length ℓ that does so, too.

Proof. Assume ω_1 is a string of length ℓ that is not n-local. By C.1, we can take $\omega_2 \neq \omega_1$ such that $\bar{\bar{\omega}}_2 = \bar{\bar{\omega}}_1 = \ell$ & $S_n(\omega_2) = S_n(\omega_1)$.

We can assume that a, b $\in$ A (even if ω_1 and ω_2 contain no a's or b's). Since $\omega_2 \neq \omega_1$ there is some position at which the two strings differ. By character-indiscernibility, then, we can take strings ω_1' and ω_2' such that $\omega_1' \neq \omega_2'$ & $|\omega_1'| = |\omega_2'| = \ell$ & $S_n(\omega_1) = S_n(\omega_2)$ and such that there is some position i_0 at which $\omega_1'(i_0) = b$ & $\omega_2'(i_0) = a$.

Now define a function $f : A \to \{a, b\}$ by

$$f(x) = \begin{cases} b, \text{ if } x = b; \\ a, \text{ otherwise.} \end{cases}$$

Define binary strings $\omega_1'' = f \circ \omega_1'$ and $\omega_2'' = f \circ \omega_2'$. Then ω_1'' and ω_2'' are distinct binary strings of length ℓ disagreeing at position i_0; and, for each binary string ω',

$$\begin{aligned}(S_n(\omega_1''))([\omega']) &= \sum_{f \circ \omega = \omega'} (S_n(\omega_1'))([\omega]) \\ &= \sum_{f \circ \omega = \omega'} (S_n(\omega_2'))([\omega]) \\ &= (S_n(\omega_2''))([\omega']).\end{aligned}$$

Therefore, ω_1'' is a binary string of length ℓ that is not n-local.

If ω_1 is a string of length ℓ that is not local, then ω_1 is not ℓ-local, by Fact 14. By the above argument, there is a binary string ω_1'' of length ℓ that is not ℓ-local. By Fact 14, ω_1'' is not local. □

The most important thing about binary sufficiency is that it provides some rationale for beginning parts of our study of observational complexity with A = {a, b}. (In this case, character-indiscernability guarantees that the truth of a statement is preserved if we replace each character mentioned by its opposite. With this in mind, we usually state our results one way (often favoring the cahracter b over the character a) and leave the statements about opposites as unstated corollaries.)

For each $k \in N$, let

$$Z^k = \{X \in R^k \mid X(i) \text{ is an integer for each } i \in k\},$$

$$Q^k = \{X \in R^k \mid X(i) \text{ is rational for each } i \in k\},$$

$$P^k = \{X \in R^k \mid X(i) > 0 \text{ for each } i \in k\},$$

$$M^k = \{X \in R^k \mid X(i) < X(j) \text{ whenever } 0 \leq i < j < k\}.$$

If ω is a binary string containing exactly k many b's, then there is a unique member X of $Z^k \cap P^k \cap M^k$ whose coordinates give the positions of ω's b's (called ω's _b-positions_); we say that X _b-represents_ ω. Of course we have similar definitions of _a-positions_ and _a-representation_.

In working with real vectors that arise as b-representations, it is occasionally convenient to have the notation V_n^k ($k, n \in N$) for the member $(1^n, 2^n, \ldots, k^n)$ of $Z^k \cap P^k$.

Before we conclude this section, we present one more quite valuable simplification, called "cancellation."

Theorem C.15 (Cancellation). Let x be any character. Strings ω_1 and ω_2 realize the same n-spectrum or spectrum iff $\omega_1 x$ and $\omega_2 x$ do.

Proof. The proof for strings realizing the same spectrum is just like that for strings realizing the same n-spectrum, so we give only the latter.

Assume $S_n(\omega_1) = S_n(\omega_2)$. Let $\omega \in A^* \;\&\; \bar{\bar{\omega}} \leq n$.

If $\omega(\bar{\bar{\omega}} - 1) \neq x$, then

$$\begin{aligned}(S_n(\omega_1 x))([\omega]) &= (S_n(\omega_1))([\omega]) \\ &= (S_n(\omega_2))([\omega]) \\ &= (S_n(\omega_2 x))([\omega]).\end{aligned}$$

If $\omega = \omega' x$ for some ω', then

$$(S_n(\omega_1 x))([\omega]) = (S_n(\omega_1))([\omega]) + (S_n(\omega_1))([\omega'])$$

$$= (S_n(\omega_2))([\omega]) + (S_n(\omega_2))([\omega'])$$
$$= (S_n(\omega_2 x))([\omega]).$$

If $\omega = x$, then

$$(S_n(\omega_1 x))([\omega]) = (S_n(\omega_1))([\omega]) + 1$$
$$= (S_n(\omega_2))([\omega]) + 1$$
$$= (S_n(\omega_2 x))([\omega]).$$

Therefore, $S_n(\omega_1 x) = S_n(\omega_2 x)$.

Conversely, assume $S_n(\omega_1 x) = S_n(\omega_2 x)$. For the sake of argument, suppose $S_n(\omega_1) \neq S_n(\omega_2)$. Let $\omega \in A^*$ ($\bar{\bar{\omega}} \leq n$) be a string of minimum length such that $(S_n(\omega_1))([\omega]) \neq (S_n(\omega_2))([\omega])$. If $\omega(\bar{\bar{\omega}} - 1) \neq x$, then

$$(S_n(\omega_1))([\omega]) = (S_n(\omega_1 x))([\omega])$$
$$= (S_n(\omega_2 x))([\omega])$$
$$= (S_n(\omega_2))([\omega]).$$

If $\omega = \omega' x$ for some ω', then $\bar{\bar{\omega}}' < \bar{\bar{\omega}}$; hence,

$$(S_n(\omega_1))([\omega]) = S_n(\omega_1 x))([\omega]) - (S_n(\omega_1))([\omega'])$$
$$= (S_n(\omega_2 x))([\omega]) - (S_n(\omega_2))([\omega'])$$
$$= (S_n(\omega_2))([\omega]).$$

If $\omega = x$, then

$$(S_n(\omega_1))([\omega]) = (S_n(\omega_1 x))([\omega]) - 1$$
$$= (S_n(\omega_2 x))([\omega]) - 1$$
$$= (S_n(\omega_2))([\omega]).$$

From this contradiction, we conclude that $S_n(\omega_1) = S_n(\omega_2)$. □

Corollary C.16. Let x be any character. Strings ω_1 and ω_2 realize the same n-spectrum or spectrum iff $x\omega_1$ and $x\omega_2$ do.

Proof. This follows from cancellation by reversibility:

$$
\begin{aligned}
S_n(\omega_1) = S_n(\omega_2) &\Leftrightarrow S_n(\hat{\omega}_1) = S_n(\hat{\omega}_2) \\
&\Leftrightarrow S_n(\hat{\omega}_1 x) = S_n(\hat{\omega}_2 x) \\
&\Leftrightarrow S_n(\hat{x\omega}_1) = S_n(\hat{x\omega}_2) \\
&\Leftrightarrow S_n(x\omega_1) = S_n(x\omega_2). \ \square
\end{aligned}
$$

Corollary C.17. Let x be any character. If ω fails to be n-local or local, then so do ωx and $x\omega$.

Proof. Immediate, by C.15 and C.16. □

Corollary C.18. If $G(\ell)$ is defined and $\ell > 0$, then there is some string ω of length ℓ such that $F(\omega) = G(\ell)$.

Remark. The significance of this result is that we need only look at strings of length ℓ to compute $G(\ell)$. (In Section E, we shall see that $G(\ell)$ is defined for every ℓ.)

Proof. Assume $\ell > 0$ & $G(\ell) = n$. By B.12 and Fact 8, $n \geq 1$. Then every string of length no greater than ℓ is n-local, but there is a string of length no greater than ℓ that is not $(n - 1)$-local. Let ω' fail to be $(n - 1)$-local, with $\bar{\bar{\omega}}' \leq \ell$. Let x be any character. By C.17, $\omega = \omega' x^{(\ell - \bar{\bar{\omega}}')}$ is a string of length ℓ that is not $(n - 1)$-local. As $\bar{\bar{\omega}} \leq \ell$, however, ω is n-local. Therefore, $F(\omega) = n$, as desired. □

Corollary C.19. If $G(\ell)$ is defined and $\ell > 0$, then there is some binary string of length ℓ such that $F(\omega) = G(\ell)$.

Remark. In fact, then, we can compute $G(\ell)$ by looking at just the (2^ℓ many) binary string of length ℓ.

Proof. By binary sufficiency, we can get the ω of the previous proof to be binary. □

Corollary C.20. Assume $G(\ell)$ is defined ($\ell > 0$). Then $G(\ell) > n$ iff there are binary strings of length ℓ realizing the same n-spectrum but beginning (or ending) with opposite characters.

Remark. This fact is a great aid in our efforts to compute values of G. To show that $G(\ell) \leq n$, it suffices to show that the first character of each binary string of length ℓ is determined by the n-spectrum realized by the string.

Proof. If there are binary strings of length ℓ realizing the same n-spectrum but beginning with opposite characters, then $G(\ell) > n$ by definition.

Conversely, assume $G(\ell) > n$. By C.19, there is a binary string ω of length ℓ with $F(\omega) = G(\ell) > n$. As ω is not n-local there is a binary string $\omega' \neq \omega$ of length ℓ with $S_n(\omega') = S_n(\omega)$, by C.1. Let i_0 be the least position at which ω' disagrees with ω, and let ω_1 and ω_2 be the binary strings of length ℓ defined as follows:

$$\omega_1(i) = \begin{cases} \omega(i + i_0), & \text{if } i + i_0 < \ell; \\ a, & \text{otherwise}; \end{cases}$$

$$\omega_2(i) = \begin{cases} \omega'(i + i_0), & \text{if } i + i_0 < \ell; \\ a, & \text{otherwise}. \end{cases}$$

By cancellation, $S_n(\omega_1) = S_n(\omega_2)$; and $\omega_1(0) = \omega(i_0) \neq \omega'(i_0) = \omega_2(0)$. □

D. Concatenation

In this section, we look at concatenations of strings that fail to be n-local. Out of this comes an easy answer to the question of how many strings can realize the same n-spectrum.

Proposition D.1. $S_n(\omega_1) = S_n(\omega_1') \;\&\; S_n(\omega_2) = S_n(\omega_2') \Rightarrow S_n(\omega_1\omega_2) = S_n(\omega_1'\omega_2')$.

Proof. Assume $S_n(\omega_1) = S_n(\omega_1') \;\&\; S_n(\omega_2) = S_n(\omega_2')$. If ω is any string (on A) of length no greater than n, then

$$\begin{aligned}(S_n(\omega_1\omega_2))\,([\omega]) &= (S_n(\omega_1))\,([\omega]) + (S_n(\omega_2))\,([\omega]) \\ &\quad + \sum_{\omega = \varphi\Psi} (S_n(\omega_1))\,([\varphi])\,(S_n(\omega_2))\,([\Psi]) \\ &= (S_n(\omega_1'))\,([\omega]) + (S_n(\omega_2'))\,([\omega]) \\ &\quad + \sum_{\omega = \varphi\Psi} (S_n(\omega_1'))\,([\varphi])\,(S_n(\omega_2'))\,([\Psi]) \\ &= S_n(\omega_1'\omega_2')\,([\omega]). \;\Box\end{aligned}$$

Corollary D.2. Let ℓ and n be given. If there are strings that fail to be n-local, then we can find ℓ many strings realizing the same n-spectrum.

Proof. Let ω fail to be n-local and let $\omega' \neq \omega$ be such that $S_n(\omega') = S_n(\omega)$. Let m be the least integer no smaller than $\log_2\ell$. In the language $\{\omega, \omega'\}^*$, then, there are $2^m \geq \ell$ many strings of length $m\bar{\omega}$. By D.1, these strings all realize the same n-spectrum. □

Notice that the proof of D.2 gives us a (perhaps poor) upper bound on how long the strings must be. For example, if there is a string of length seven that is not 3-local, then we can find ℓ many strings of length 7m that realize the same 3-spectrum, where m is the least integer no smaller than $\log_2\ell$.

Example. It turns out that S_3(baaabba) = S_3(abbaaab). (See the proof of Corollary G.7, below.) Here, then, are eight strings that realize the same 3-spectrum.

baaabbabaaabbabaaabba,

baaabbabaaabbaabbaaab,

baaabbaabbaaabbaaabba,

baaabbaabbaaababbaaab,

abbaaabbaaabbabaaabba,

abbaaabbaaabbaabbaaab,

abbaaababbaaabbaaabba,

abbaaababbaaababbaaab.

Also, observe that we can use simple cancellation (C.15, C.16) to pad our strings; thus we can get ℓ many strings that realize the same n-spectrum and that conform to some set compositional ratios. (Zero ratios can only be approximated, of course; for we are, presumably, stuck with the original strings' characters.) In Section G, we demonstrate an entirely different procedure for finding binary strings realizing the same 3-spectrum. The procedure there is much more difficult and does not give such a clear bound on how long the strings must be, but it is a much stronger result in one way: each of the strings obtained contains just four b's!

We give just one more sample application of D.1.

Corollary D.3. If there is any string that fails to be n-local, then there is a palindrome (i.e., a string that is its own reverse) that fails to be n-local.

<u>Proof</u>. Assume $F(\omega) > n$, and take $\omega' \neq \omega$ so that $S_n(\omega') = S_n(\omega)$. By reversibility, $S_n(\hat{\omega}') = S_n(\hat{\omega})$; so, by D.1, $S_n(\omega'\hat{\omega}') = S_n(\omega\hat{\omega})$. Both $\omega\hat{\omega}$ and $\omega'\hat{\omega}'$, then, are palindromes that fail to be n-local. □

E. Upper Bounds on F and G

In this section we attempt to get an idea of the nature of F and G by finding some upper bounds. Our first result gives a natural upper bound for F at each point (string) in its domain.

Proposition E.1. Every string of length ℓ is ℓ-local.

Proof. Suppose some string ω_1 of length ℓ is not ℓ-local. By C.1, there is a string $\omega_2 \neq \omega_1$ of length ℓ such that $S_\ell(\omega_2) = S_\ell(\omega_1)$. As $(S_\ell(\omega_1))([\omega_1]) \neq 0$, the event $[\omega_1]$ must also occur in ω_2. But $[\omega_1]$ can only occur is a string of length ℓ if that string is ω_1; therefore, $\omega_2 = \omega_1$, a contradiction. □

Corollary E.2. Every string is local.

Proof. Immediate, by E.1; for every string has a length. □

Corollary E.3. Every finite language is local.

Proof. Let L be a finite language, and let ℓ be the length of the longest string in L. By E.1 and Fact 5, every string in L is ℓ-local. By Fact 9, then, L is ℓ-local, hence local. □

Corollary E.4. F is total on strings and finite languages.

Proof. Let x be any string or finite language. By E.2 and E.3, $\{n \mid x$ is n-local$\} \neq \emptyset$. As N is well-ordered, there is a least n such that x is n-local. □

Corollary E.5. $F(\omega) \leq \bar{\bar{\omega}}$ for each string ω.

Proof. Immediate, by E.1. and E.4. □

Corollary E.6. We can effectively compute $F(\omega)$.

Proof. Let ω be any given string. There are only finitely many rearrangements of ω, so we can make a complete list of them. Since each rearrangement of ω has length $\bar{\bar{\omega}}$ and there are only finitely many complete events of extent no greater than $\bar{\bar{\omega}}$, we can easily determine the entire n-spectrum (any n) realized by each rearrangement. If we do this for successive values of n until we get one for which none of the other rearrangements of ω realizes the same n-spectrum as ω, then that value is $F(\omega)$, by C.1. (By E.4, we do eventually find such an n.) □

Example. Consider the string ω = abba. The other rearrangements of ω are aabb, abab, baab, baba, bbaa. All of these strings realize the 0-spectrum $S_0(\omega) = \emptyset$ and the 1-spectrum $\{([a], 2), ([b], 2)\}$, of course. Next, we compute 2-spectra:

$$S_2(\omega) = S_1(\omega) \cup \{([aa], 1), ([bb], 1), ([ab], 2)\ ([ba], 2)\};$$
$$S_2(\text{aabb}) = S_1(\omega) \cup \{([aa], 1), ([bb], 1), ([ab, 4), ([ba], 0)\};$$
$$S_2(\text{abab}) = S_1(\omega) \cup \{([aa], 1), ([bb], 1), ([ab], 3), ([ba, 1)\};$$
$$S_2(\text{baab}) = S_1(\omega) \cup \{([aa], 1), ([bb], 1), ([ab], 2), ([ba], 2)\};$$
$$S_2(\text{baba}) = S_1(\omega) \cup \{([aa], 1), ([bb], 1), ([ab], 1), ([ba], 3)\};$$
$$S_2(\text{bbaa}) = S_1(\omega) \cup \{([aa], 1), ([bb], 1), ([ab], 0), ([ba], 4)\}.$$

Only baab realizes $S_2(\text{abba})$; when we go on to consider 3-spectra, then, Fact 3 guarantees that we need only consider this particular rearrangement of ω. Finally, we compute

$$S_3(\omega) = S_2(\omega) \cup \{([aaa], 0), ([bbb], 0), ([aab], 0), ([aba], 2), ([baa], 0), ([bba], 1), ([bab], 0), ([abb], 1)\};$$

$$S_3(\text{baab}) = S_2(\omega) \cup \{([aaa], 0), ([bbb], 0), ([aab], 1), ([aba], 0), ([baa], 1), ([bba], 0), ([bab], 2), ([abb], 0)\}.$$

These 3-spectra differ, so $F(\omega) = 3$.

Corollary E.7. G is total.

Proof. There are only finitely many (2^{ℓ}) binary strings of length ℓ. By C.19, $G(\ell)$ is just the maximum value of F on these strings. (F is defined on these strings, by E.4.) □

Corollary E.8. For each $\ell > 0$ there is some string of length ℓ such that $F(\omega) = G(\ell)$.

Proof. Given E.7, this is just a restatement of C.19. □

Corollary E.9. $G(\ell) \leq \ell$ for each $\ell \in N$.

Proof. Immediate, as $G(0) = 0 \leq 0$ and by C.19 and E.5. □

Corollary E.10. We can effectively compute $G(\ell)$.

Proof. Immediate, by C.19 and E.6; for we can effectively list the 2^{ℓ} many binary strings of length ℓ and effectively compute F. □

Example. Consider $\ell = 4$. The binary strings of length 4 are aaaa, bbbb, the rearrangements of abba, the rearrangements of abbb, and the rearrangements of baaa. We know, by B.10 that F(aaaa) = F(bbbb) = 1. Computations show that, if ω is a rearrangement of either abbb or baaa, then $F(\omega) = 2$. The computation of our previous example showed that F(abba) = F(baab) = 3 and that $F(\omega) = 2$ for any other rearrangement ω of abba. Therefore, $G(4) = 3$.

For large ℓ, there are very many ($\bar{\bar{A}}^{\ell}$) strings of length ℓ and often very many rearrangements of a given string of length ℓ; so the above methods of computing F and G are very gross. Later on we demonstrate some shortcuts.

Now we ask whether our bound on F and G are good ones. First we ask whether the bounds are assumed anywhere. Our next result and B.12 give affirmative, though trivial, answers.

Proposition E.11. If $\bar{\bar{\omega}} = 1$, then $F(\omega) = 1$.

Proof. Immediate, by B.10; for any string of length one is constant. □

In other cases, however, it turns out that we can get considerably better upper bounds. To do so we must work a little harder.

Theorem E.12. If $n > \frac{\ell}{2}$, then $G(\ell) \le n$.

Proof. If $n \ge \ell$, then E.9 tells us that $G(\ell) \le \ell \le n$.

Suppose $\ell > n > \frac{\ell}{2}$ and $G(\ell) \nleq n$. By C.20 and E.7, there are binary strings ω_1 and ω_2 of length ℓ, with $\omega_1(0) = b$ & $\omega_2(0) = a$ & $S_n(\omega_1) = S_n(\omega_2)$.

The number of n-reports on a string of length ℓ is $\binom{\ell}{n} = \frac{\ell!}{n!\,(\ell-n)!}$.

Now, the number of these that start at the beginning of the string is $\binom{\ell-1}{n-1} = \frac{(\ell-1)!}{(n-1)!\;(\ell-n)!}$. Therefore,

$$\sum_{\substack{\bar{\bar{\omega}}=n \\ \omega(0)=b}} (S_n(\omega_1))\;([\omega]) \ge \frac{(\ell-1)!}{(n-1)!\;(\ell-n)!}$$

$$\&\quad \sum_{\substack{\bar{\bar{\omega}}=n \\ \omega(0)=a}} (S_n(\omega_2))\;([\omega]) \ge \frac{(\ell-1)!}{(n-1)!\;(\ell-n)!}\,.$$

As $\sum_{\bar{\bar{\omega}}=n} (S_n(\omega_1))\;([\omega]) = \sum_{\bar{\bar{\omega}}=n} (S_n(\omega_2))\;([\omega]) = \frac{\ell!}{n!\;(\ell-n)!}$, then,

$\frac{2(\ell-1)!}{(n-1)!\;(\ell-n)!} \le \frac{\ell!}{n!\;(\ell-n)!}$. But this simplifies to $n \le \frac{\ell}{2}$, a contradiction; hence, $\ell > n > \frac{\ell}{2} \Rightarrow G(\ell) \le n$. □

Beyond $\ell = 1$, then, we have improvement of the bound on G:

Corollary E.13. $G(\ell) < \ell$ iff $\ell > 2$.

Proof. By Fact 7 and Corollary B.12, $G(0) = 0$ & $G(1) = 1$; and an easy computation shows that $G(2) = 2$.

Conversely, if $\ell > 2$ and n is the least integer greater than $\frac{\ell}{2}$, then $n < \ell$; and $G(\ell) \leq n$, by E.12. □

Corollary E.14. $F(\omega) < \bar{\bar{\omega}}$ unless $\bar{\bar{\omega}} = 1$ or $\bar{\bar{\omega}} = 2$.

Proof. This follows immediately from E.13 and the definition of G. □

We cannot, in general, strengthen our result to get either $n > \frac{\ell}{2} \Rightarrow G(\ell) < n$ or $n = \frac{\ell}{2} \Rightarrow G(\ell) \leq n$. For example,

$$F(ab) = F(ba) = 2 > 1,$$

$$F(abba) = F(baab) = 3 > 2.$$

For $\ell = 2n > 4$, however, we do have $G(\ell) \leq n$.

Theorem E.15. If $\ell = 2n > 4$, then $G(\ell) \leq n$.

Proof. Suppose $\ell = 2n > 4$ and $G(\ell) > n$. By C.20 and E.7 there are binary strings ω_1 and ω_2 of length ℓ, with $\omega_1(0) = b$ & $\omega_2(0) = a$ & $S_n(\omega_1) = S_n(\omega_2)$. As in the proof of E.12,

$$\sum_{\bar{\bar{\omega}}=n} (S_n(\omega_1))([\omega]) = \sum_{\bar{\bar{\omega}}=n} (S_n(\omega_2))([\omega]) = \frac{\ell!}{n!\ (\ell-n)!}$$

$$\& \sum_{\substack{\bar{\bar{\omega}}=n \\ \omega(0)=b}} (S_n(\omega_1))([\omega]) \geq \frac{(\ell-1)!}{(n-1)!\ (\ell-n)!}$$

$$\& \sum_{\substack{\bar{\bar{\omega}}=n \\ \omega(0)=a}} (S_n(\omega_2))([\omega]) \geq \frac{(\ell-1)!}{(n-1)!\ (\ell-n)!}.$$

$$\text{If} \sum_{\substack{\bar{\bar{\omega}}=n \\ \omega(0)=b}} (S_n(\omega_1))([\omega]) > \frac{(\ell-1)!}{(n-1)!\ (\ell-n)!}$$

$$\text{or} \sum_{\substack{\bar{\bar{\omega}}=n \\ \omega(0)=a}} (S_n(\omega_2))([\omega]) > \frac{(\ell-1)!}{(n-1)!\ (\ell-n)!},$$

then $\frac{2(\ell-1)!}{(n-1)!\ (\ell-n)!} < \frac{\ell!}{n!\ (\ell-n)!}$, which simplifies to $\ell > 2n$, a contradiction. Therefore, both inequalities are actually equalities. From this we can conclude two things:

(i) An n-report on ω_1 is of the form $[\omega]$, with $\omega(0) = b$, iff it starts at the beginning of the string.

(ii) An n-report on ω_2 is of the form $[\omega]$, with $\omega(0) = a$, iff it starts at the beginning of the string.

Hence, we must have $\omega_1(i) = a$ & $\omega_2(i) = b$ for $1 \leq i \leq \ell-n$. Since $\ell = 2n$, then, ω_1 contains at least $n = \frac{\ell}{2}$ many a's, and ω_2 contains at least $n = \frac{\ell}{2}$ many b's. By B.6 and Fact 3, ω_2 is a rearrangement of ω_1, however; so each of ω_1 and ω_2 is a rearrangement of $b^n a^n$. Therefore, $\omega_1 = ba^n b^{n-1}$ & $\omega_2 = ab^n a^{n-1}$. We know, then, that $(S_3(\omega_1))([bab]) \neq 0 = (S_3(\omega_2))([bab])$, so that $S_3(\omega_1) \neq S_3(\omega_2)$. But $\ell = 2n > 4 \Rightarrow n \geq 3$, so that $S_3(\omega_1) = S_3(\omega_2)$, a contradiction. Therefore, $\ell = 2n > 4 \Rightarrow G(\ell) \leq n$. □

Even this result can "fail" for ℓ odd, however. For example, it can be verified that F(baaabba) = F(abbaaab) $= 4 > \frac{7}{2}$. So one might guess that we are just about as low as we can go with our bounds. In particular, one might conjecture that we never get $G(\ell) < \frac{\ell}{2}$. One more empirical work will refute this conjecture, however ; so perhaps there are still lower upper bounds on F and G that can be expressed.

F. 2-events

Seeing no obvious general way to continue our study of F and G, we turn now to more empirical study.

We already know from our examples (see page 49) that there are strings which are not 2-local, but it remains to nicely characterize these strings. Now we restrict our attention to binary strings and proceed to do so. (Recall the notation and definitions introduced on pages 40-41.)

Theorem F.1. Binary strings realize the same 2-spectrum iff they are rearrangements with the same average b-position.

Remarks. (i) If $X, Y \in R^k$ b-represent binary strings ω_1 and ω_2, respectively, of the same length, then F.1 says that $S_2(\omega_1) = S_2(\omega_2) \Leftrightarrow \Sigma x_i = \Sigma y_i$.

(ii) A consequence of F.1 is that, given the length of a binary string b-represented by X and given the number of b's the string contains, the single number Σx_i completes the characterization of the 2-spectrum realized by the string.

Proof. Let ω_1 and ω_2 be binary strings.

If ω_1 and ω_2 are not rearrangements, then neither do they realize the same 2-spectrum (by B.6 and Fact 3), nor are they rearrangements with the same average b-position.

Assume, now, that ω_1 and ω_2 are rearrangements, and let $X, Y \in R^k$ b-represent ω_1 and ω_2 respectively. Let $\ell = \bar{\bar{\omega}}_1 = \bar{\bar{\omega}}_2$. Since ω_1 and ω_2 are rearrangements,

$$S_1(\omega_1) = S_1(\omega_2)$$

$$\& \quad (S_2(\omega_1))([bb]) = (S_2(\omega_2))([bb])$$

$$\& \quad (S_2(\omega_1))\,([aa]) = (S_2(\omega_2))\,([aa]);$$

so $S_2(\omega_1) = S_2(\omega_2) \Leftrightarrow (S_2(\omega_1))\,([ba]) = (S_2(\omega_2))\,([ba])$

$$\& \ (S_2(\omega_1))\,([ab]) = (S_2(\omega_2))\,([ab]).$$

Now, if f(i) is the cardinality of $\{j > x_i - 1 \mid \omega_1(j) = a\}$, then

$$(S_2(\omega_1))\,([ba]) = \sum_{i=1}^{k} f(i).$$

As $f(i) = (\ell-k) - (x_i - i)$,

$$(S_2(\omega_1))\,([ba]) = \sum_{i=1}^{k} [(\ell-k) - (x_i - i)]$$

$$= k(\ell-k) - \sum_{i=1}^{k} x_i + \sum_{i=1}^{k} i.$$

Similarly, $(S_2(\omega_2))\,([ba]) = k(\ell-k) - \Sigma y_i + \Sigma i.$

Also, $(S_2(\omega_1))\,([ab]) = \sum_{i=1}^{k} (x_i - i) = \sum_{i=1}^{k} x_i - \sum_{i=1}^{k} i;$

similarly, $(S_2\ (\omega_2))\,([ab]) = \Sigma y_i - \Sigma i.$

Clearly, then,

$$(S_2(\omega_1))\,([ba]) = (S_2(\omega_2))\,([ba])$$

$$\& \ (S_2(\omega_1))\,([ab]) = (S_2(\omega_2))\,([ab])$$

$$\Leftrightarrow \ \Sigma x_i = \Sigma y_i.$$

Therefore, $S_2(\omega_1) = S_2(\omega_2) \Leftrightarrow \Sigma x_i = \Sigma y_i$. □

Corollary F.2. Any binary string containing fewer than two b's is 2-local.

Proof. If a string is constantly a, it is 1-local (by B.10); and distinct rearrangements of a string containing exactly one b have different average b-positions. The result follows by F.1. □

Now we record a result we already stumbled on in the example on page 49.

Corollary F.3. There are strings which are not 2-local.

Proof. By F.1, S_2(abba) = S_2(baab) and S_2(ababa) = S_2(baaab), for example. □

We can in fact strengthen F.3 to get another corollary.

Corollary F.4. The shortest string that is not 2-local has length four.

Proof. We have already seen (in F.3) that abba is a string of length four that is not 2-local. By F.2, any binary string containing fewer than two b's is 2-local. By character-indiscernibility we know also that any binary string containing fewer than two a's is 2-local. Therefore, any binary string that is not 2-local contains at least two b's and at least two a's. By binary sufficiency it follows that any string that is not 2-local has length at least four. □

Finally, we may as well observe that we have one more corollary.

Corollary F.5. There are strings of observational complexity three.

Proof. Just take abba. (See the example on page 49.) □

G. 3-events

It is much more difficult to find strings that are not 3-local than it is to find strings that are not 2-local. By E.15 in fact we know that only strings of length greater than six can fail to be 3-local. We could begin to look at the binary strings of length seven, but there are $2^7 = 128$ of them. Reversibility, character-indiscernibility, and cancellation would limit the size of our search, and F.1 could make it easier (by Fact 3); but it is still more instructive to look for a general algebraic characterization of binary strings that are not 3-local. Our next result does this in the same sense that F.1 gives a general algebraic characterization of binary strings that are not 2-local.

Theorem G.1. Let ω_1 and ω_2 be binary strings realizing the same 2-spectrum, and let X and Y b-represent ω_1 and ω_2, respectively. Then $S_3(\omega_1) = S_3(\omega_2)$ iff $\Sigma ix_i = \Sigma iy_i$ & $\Sigma x_i^2 = \Sigma y_i^2$.

Proof. As $S_2(\omega_1) = S_2(\omega_2)$, ω_1 and ω_2 are rearrangements with $\Sigma x_i = \Sigma y_i$, by F.1. Let $\ell = \bar{\bar{\omega}}_1 = \bar{\bar{\omega}}_2$, and let $k = \bar{\bar{X}} = \bar{\bar{Y}}$. As $S_2(\omega_1) = S_2(\omega_2)$, $S_3(\omega_1) = S_3(\omega_2)$ iff $S_3(\omega_1)$ and $S_3(\omega_2)$ agree on all 3-events. The strings ω_1 and ω_2 are rearrangements, so $S_3(\omega_1)$ and $S_3(\omega_2)$ agree on [bbb] and [aaa]. The only other 3-events are [abb], [aab], [bba], [baa], [aba], and [bab].

For each i (i = 1, 2, ..., k), let f(i), f′(i), g(i) g′(i) be the respective cardinalities of

$$\{j < x_i - 1 \mid \omega_1(j) = b\},$$
$$\{j < x_i - 1 \mid \omega_1(j) = a\},$$
$$\{j > x_i - 1 \mid \omega_1(j) = b\},$$
$$\{j > x_i - 1 \mid \omega_1(j) = a\}.$$

Explicitly, then,

$$f(i) = i - 1,$$
$$f'(i) = x_i - 1,$$
$$g(i) = k - i,$$
$$g'(i) = (\ell - k) - (x_i - i).$$

Now, $(S_3(\omega_1))\ ([abb]) = \sum_{i=1}^{k} f'(i)\ g(i)$

$$= \sum_{i=1}^{k} (x_i - i)\ (k - i)$$

$$= k \sum_{i=1}^{k} x_i - \sum_{i=1}^{k} ix_i - k \sum_{i=1}^{k} i + \sum_{i=1}^{k} i^2.$$

Similarly, $(S_3(\omega_2))\ ([abb]) = k\Sigma y_i - \Sigma iy_i - k\Sigma i + \Sigma i^2$.

Therefore,

$$(S_3(\omega_1))\ ([abb]) = (S_3(\omega_2))\ ([abb])$$
$$\Leftrightarrow k\Sigma x_i - \Sigma ix_i = k\Sigma y_i - \Sigma iy_i.$$

Continuing our calculations, we get

$$(S_3(\omega_1))\ ([aab]) = \sum_{i=1}^{k} \binom{f'(i)}{2} = \sum_{i=1}^{k} \frac{(x_i - i)\ (x_i - i - 1)}{2!}$$

$$= \frac{1}{2}\sum_{i=1}^{k} x_i^{\ 2} - \frac{1}{2}\sum_{i=1}^{k} ix_i - \frac{1}{2}\sum_{i=1}^{k} x_i - \frac{1}{2}\sum_{i=1}^{k} ix_i + \frac{1}{2}\sum_{i=1}^{k} i^2 + \frac{1}{2}\sum_{i=1}^{k} i$$

$$= \frac{1}{2}\sum_{i=1}^{k} x_i^{\ 2} - \sum_{i=1}^{k} ix_i - \frac{1}{2}\sum_{i=1}^{k} x_i + \frac{1}{2}\sum_{i=1}^{k} i^2 + \frac{1}{2}\sum_{i=1}^{k} i,$$

and $(S_3(\omega_2))\ ([aab]) = \frac{1}{2}\ \Sigma y_i^{\ 2} - \Sigma iy_i - \frac{1}{2}\ \Sigma y_i + \frac{1}{2}\ \Sigma i^2 + \frac{1}{2}\ \Sigma i.$

Therefore,

$$(S_3(\omega_1))\ ([aab]) = (S_3(\omega_2))\ ([aab])$$

$$\Leftrightarrow \frac{1}{2}\ \Sigma x_i^{\ 2} - \Sigma ix_i - \frac{1}{2}\ \Sigma x_i = \frac{1}{2}\ \Sigma y_i^{\ 2} - \Sigma iy_i - \frac{1}{2}\ \Sigma y_i.$$

Accumulating our findings, we get

$$S_3(\omega_1) = S_3(\omega_2)$$

$$\Leftrightarrow \quad k\Sigma x_i - \Sigma i x_i = k\Sigma y_i - \Sigma i y_i$$

$$\& \ \frac{1}{2}\Sigma x_i^{\,2} - \Sigma i x_i - \frac{1}{2}\Sigma x_i = \frac{1}{2}\Sigma y_i^{\,2} - \Sigma i y_i - \frac{1}{2}\Sigma y_i$$

& $S_3(\omega_1)$, $S_3(\omega_2)$ agree on [bba], [baa], [aba], [bab].

As $\Sigma x_i = \Sigma y_i$, this simplifies to

$$S_3(\omega_1) = S_3(\omega_2) \Leftrightarrow$$

$$\Sigma i x_i = \Sigma i y_i \ \& \ \Sigma x_i^{\,2} = \Sigma y_i^{\,2}$$

& $S_3(\omega_1)$, $S_3(\omega_2)$ agree on [bba], [baa], [aba], [bab].

It will suffice, then, to show that

$\Sigma i x_i = \Sigma i y_i \ \& \ \Sigma x_i^{\,2} = \Sigma y_i^{\,2} \Rightarrow S_3(\omega_1), S_3(\omega_2)$ agree on [bba], [baa], [aba], [bab]. So assume $\Sigma i x_i = \Sigma i y_i \ \& \ \Sigma x_i^{\,2} = \Sigma y_i^{\,2}$. Clearly, then, if $P(i,x)$ is any polynomial in i and x having degree at most two, then $\sum_{i=1}^{k} P(i,x_i) = \sum_{i=1}^{k} P(i,y_i)$; we use this fact to show that $S_3(\omega_1)$, $S_3(\omega_2)$ agree on [bba], [baa], [aba], [bab].

$$(S_3(\omega_1))\ ([bba]) = \sum_{i=1}^{k} f(i)\ g'(i)$$

$$= \sum_{i=1}^{k} (i-1)\ [(\ell-k) - (x_i-i)]$$

$$= \sum_{i=1}^{k} (i-1)\ [(\ell-k) - (y_i-i)]$$

$$= (S_3(\omega_2))\ ([bba]).$$

$$(S_3(\omega_1))\ ([baa]) = \sum_{i=1}^{k} \binom{g'(i)}{2}$$

$$= \frac{1}{2}\sum_{i=1}^{k} [(\ell-k) - (x_i-i)] [(\ell-k) - (x_i-i) - 1]$$

$$= \frac{1}{2}\sum_{i=1}^{k} [(\ell-k) - (y_i-i)] [(\ell-k) - (y_i-i) - 1]$$

$$= (S_3(\omega_2)) ([baa]).$$

$$(S_3(\omega_1)) ([aba]) = \sum_{i=1}^{k} f'(i) g'(i)$$

$$= \sum_{i=1}^{k} (x_i-i) [(\ell-k) - (x_i-i)]$$

$$= \sum_{i=1}^{k} (y_i-i) [(\ell-k) - (y_i-i)]$$

$$= (S_3(\omega_2)) ([aba]).$$

$$(S_3(\omega_1)) ([bab]) = \sum_{i=1}^{k} \sum_{j=i}^{k} [f'(j) - f'(i)]$$

$$= \sum_{i=1}^{k} \sum_{j=i}^{k} [(x_j-j) - (x_i-i)]$$

$$= \sum_{i=1}^{k} \sum_{j=i}^{k} (x_j-j) - \sum_{i=1}^{k} \sum_{j=i}^{k} (x_i-i)$$

$$= \sum_{i=1}^{k} i(x_i-i) - \sum_{i=1}^{k} (x_i-i) (k-i+1)$$

$$= \sum_{i=1}^{k} i(y_i-i) - \sum_{i=1}^{k} (y_i-i) (k-i+1)$$

$$= (S_3(\omega_2)) ([bab]). \square$$

Corollary G.2. Let ω_1 and ω_2 be binary strings b-represented by X and Y, respectively. Then $S_3(\omega_1) = S_3(\omega_2)$ iff ω_2 is a rearrangement of ω_1 and

$$\Sigma x_i = \Sigma y_i$$
$$\& \ \Sigma ix_i = \Sigma iy_i$$
$$\& \ \Sigma x_i^2 = \Sigma y_i^2.$$

Remark. A significant consequence of this corollary is that, given the length of a binary string b-represented by X and given the number of b's the string contains, just three additional numbers (Σx_i, Σix_i, and Σx_i^2) are needed to entirely characterize the 3-spectrum realized by the string. (We shall see shortly that these numbers alone need not pin down the string itself, however; i.e., we will display some strings that are not 3-local.)

Proof. By G.1,

$$S_2(\omega_1) = S_2(\omega_2) \ \& \ S_3(\omega_1) = S_3(\omega_2) \Leftrightarrow$$
$$S_2(\omega_1) = S_2(\omega_2) \ \& \ \Sigma ix_i = \Sigma iy_i \ \& \ \Sigma x_i^2 = \Sigma y_i^2.$$

By Fact 3 and Theorem F.1, we get the desired result from this. □

Now we use G.2 to seek strings which are not 3-local. First, we try strings containing just two b's.

Lemma G.3. There do not exist distinct real vectors X, Y $\in R^2$ such that $\Sigma x_i = \Sigma y_i$ & Σy_i & $\Sigma ix_i = \Sigma iy_i$.

Proof. Suppose $x_1 + x_2 = Y_1 + y_2$

$$\& \ x_1 + 2x_2 = y_1 + 2y_2.$$

Subtracting, we get $x_2 = y_2$, so that we must also have $x_1 = y_1$. □

Corollary G.4. Any binary string containing exactly two b's is 3-local.

Proof. Suppose ω contains exactly two b's and $S_3(\omega') = S_3(\omega)$. By Fact 3 and B.6, ω' is a rearrangement of ω. Let $X, Y \in R^2$ b-represent ω and ω', respectively. By G.2, $\Sigma x_i = \Sigma y_i$ & $\Sigma ix_i = \Sigma iy_i$. By Lemma G.3, then, $X = Y$. As ω' is a rearrangement of ω with the same b-representation, we must have $\omega' = \omega$. □

Corollary G.5. Any binary string containing no more than two b's is 3-local.

Proof. Immediate, by F.2, Fact 5, and G.4. □

We know now that we must investigate binary strings containing at least three b's (and at least three a's, by character-indiscernibility) if we hope to get distinct strings realizing the same 3-spectrum. By E.15, only strings longer than six can fail to be 3-local. The best we can hope for, then, is a rearrangement of b^3a^4 that is not 3-local.

Lemma G.6. There exist distinct real vectors $X, Y \in Z^3 \cap M^3 \cap P^3$ such that

$$\Sigma x_i = \Sigma y_i$$
$$\& \quad \Sigma ix_i = \Sigma iy_i$$
$$\& \quad \Sigma x_i^2 = \Sigma y_i^2.$$

Proof. Let $X = (1, 5, 6)$, $Y = (2, 3, 7)$. Then

$$\Sigma x_i = 12 = \Sigma y_i$$
$$\& \quad \Sigma ix_i = 29 = \Sigma iy_i$$
$$\& \quad \Sigma x_i^2 = 62 = \Sigma y_i^2.$$

(Other examples: (1, 6, 8), (2, 4 9);
(4, 8, 9), (5, 6, 10).) □

Corollary G.7. There are strings that are not 3-local.

Proof. Let ω_1 = baaabba, ω_2 = abbaaab. As ω_1 and ω_2 are rearrangements b-represented by (1, 5, 6) and (2, 3, 7) (see the previous proof), we must have $S_3(\omega_1) = S_3(\omega_2)$, by G.2. □

Corollary G.8. $G(7) > 3$.

Proof. We just saw that F(baaabba) > 3. □

Remark. This incidental result adds to our knowledge of the function G. So far, now, we have the following information:

$G(0) = 0$ (by Fact 7),

$G(1) = 1$ (by B.12),

$G(2) = 2$ (by B.12 and E.9),

$G(3) = 2$ (by Fact 8 and E.12),

$G(4) = 3$ (by the example on page 50),

$G(5) = 3$ (by Fact 8 and E.12),

$G(6) = 3$ (by Fact 8 and E.15),

$G(7) = 4$ (by G.8 and E.12),

$G(8) = 4$ (by Fact 8 and E.15),

$G(9) = 4$ or 5 (by Fact 8 and E.12),

$G(10) = 4$ or 5 (by Fact 8 and E.15),

etc.

Lemma G.9. Let $X, Y \in R^k$, $t \in R$.

If $\Sigma x_i = \Sigma y_i$, then

$$\Sigma {x_i}^2 = \Sigma {y_i}^2 \Leftrightarrow \Sigma(x_i - t)^2 = \Sigma(y_i - t)^2.$$

Proof. Assume $\Sigma x_i = \Sigma y_i$.

As $\Sigma(x_i - t)^2 = \Sigma {x_i}^2 - 2t\,\Sigma x_i + \Sigma t^2$

$= \Sigma {x_i}^2 - 2t\,\Sigma x_i + kt^2$,

$\Sigma(x_i - t)^2 = \Sigma(y_i - t)^2$

$\Leftrightarrow \Sigma {x_i}^2 + kt^2 = \Sigma {y_i}^2 + kt^2$

$\Leftrightarrow \Sigma x_i^2 = \Sigma y_i^2$. □

Lemma G.10. Let X, Y $\in R^3$. If $\Sigma x_i = \Sigma y_i$,

then

$$\Sigma i x_i = \Sigma i y_i \Leftrightarrow x_3 - x_1 = y_3 - y_1.$$

Proof. Assume $\Sigma x_i = \Sigma y_i$. As $x_3 - x_1 = \Sigma i x_i - 2\Sigma x_i$,

$$x_3 - x_1 = y_3 - y_1 \Leftrightarrow \Sigma i x_i - 2\Sigma x_i = \Sigma i y_i - 2\Sigma y_i$$

$$\Leftrightarrow \Sigma i x_i = \Sigma i y_i. \ \square$$

Lemma G.11. Let X $\in R^3$, and let $\Sigma x_i = 3m$.

Then $x_2 - x_1 \neq x_3 - x_2$ iff there is some Y $\in R^3$ different from X such that

$$\Sigma x_i = \Sigma y_i$$

$$\& \quad \Sigma i x_i = \Sigma i y_i$$

$$\& \quad \Sigma x_i^2 = \Sigma y_i^2.$$

Furthermore,

$$\{Y \in R^3 \mid \Sigma x_i = \Sigma y_i \ \& \ \Sigma i x_i = \Sigma i y_i \ \& \ \Sigma x_i^2 = \Sigma y_i^2\}$$

$$= \{X, (2m - x_3, 2m - x_2, 2m - x_1)\}.$$

Remark. In other words, the only "candidate" for Y $\neq$ X is $(2m-x_3, 2m-x_2, 2m-x_1)$; and it succeeds iff $x_2 - x_1 \neq x_3 - x_2$.

Proof. Suppose Y $\in R^3$ and

$$\Sigma x_i = \Sigma y_i$$

$$\& \quad \Sigma i x_i = \Sigma i y_i$$

$$\& \quad \Sigma x_i^2 = \Sigma y_i^2.$$

By Lemma G.10, $x_3 - x_1 = y_3 - y_1$; so $y_1 - x_1 = y_3 - x_3$. Let $d = y_1 - x_1 = y_3 - x_3$. Since $\Sigma x_i = \Sigma y_i$, we must have $y_2 = x_2 - 2d$. (Then $Y = X + d(1, -2, 1)$ is different from X iff $d \neq 0$.) Therefore,

$$\Sigma x_i^2 = \Sigma y_i^2$$

$$= (x_1+d)^2 + (x_2-2d)^2 + (x_3+d)^2$$
$$= \Sigma x_i^2 + 6d^2 + 2d\ (x_1-2x_2+x_3),$$

so that $6d^2 + 2d\ (x_1-2x_2+x_3) = 0$.

To have $Y \neq X$, we must have $d \neq 0$; so we must have $d = \frac{1}{3}\ (2x_2-x_1-x_3)$.

As

$$X + \tfrac{1}{3}(2x_2-x_1-x_3)\ (1, -2, 1)$$
$$= X + (x_2-m)\ (1, -2, 1)$$
$$= (x_1 + x_2 - m,\ 2m - x_2,\ x_2 + x_3 - m)$$
$$= (2m - x_3,\ 2m - x_2,\ 2m - x_1),$$

$\{Y \in R^3 \mid \Sigma x_i = \Sigma y_i\ \&\ \Sigma i x_i = \Sigma i y_i\ \&\ \Sigma x_i^2 = \Sigma y_i^2\}$

$\subseteq \{X, (2m - x_3, 2m - x_2, 2m - x_1)\}$.

Now, let $Y = (2m - x_3, 2m - x_2, 2m - x_1)$. Then $\Sigma y_i = 6m - \Sigma x_i = 2\Sigma x_i - \Sigma x_i = \Sigma x_i$. Since $y_3 - y_1 = (2m - x_1) - (2m - x_3) = x_3 - x_1$, Lemma G.10 guarantees that $\Sigma i y_i = \Sigma i x_i$. Since $\Sigma(y_i - m)^2 = \Sigma(m - x_i)^2 = \Sigma(x_i - m)^2$, Lemma G.9 guarantees that $\Sigma y_i^2 = \Sigma x_i^2$. Therefore,

$$\{Y \in R^3 \mid \Sigma x_i = \Sigma y_i\ \&\ \Sigma i x_i = \Sigma i y_i\ \&\ \Sigma x_i^2 = \Sigma y_i^2\}$$
$$\supseteq \{X, (2m - x_3, 2m - x_2, 2m - x_1)\}.$$

It remains only to show that $x_2 - m = 0$ iff $x_2 - x_1 = x_3 - x_2$:

$$x_2 - m = 0 \Leftrightarrow 3x_2 = 3m$$
$$\Leftrightarrow 3x_2 = x_1 + x_2 + x_3$$
$$\Leftrightarrow x_2 - x_1 = x_3 - x_2. \ \square$$

Corollary G.12. If a binary string ω contains exactly three b's, then there is at most one other binary string that realizes the same 3-spectrum as ω. If X b-represents ω and $\Sigma x_i = 3m$, then $Y = (2m - x_3, 2m - x_2, 2m - x_1)$ b-represents that other string; in fact, if Y b-represents any string of length $\bar{\bar{\omega}}$, then that string realizes $S_3(\omega)$.

Proof. Immediate from G.11, by G.2. □

Another consequence of G.2 is the following sufficient condition for 3-localness.

Proposition G.13. Let $X \in Z^k \cap M^k \cap P^k$ b-represent the binary string ω. If X is a linear combination of V_0^k and V_1^k, then ω is 3-local. (Recall the definitions of V_0^k and V_1^k, page 41.)

Proof. Assume X is a linear combination of (the linearly independent real vectors) V_0^k and V_1^k. Let θ_0 and θ_1 be the respective angles that X makes with V_0^k and V_1^k. (Angles are taken in the interval $[0, \pi)$.) It is a simple fact of euclidean geometry that there is no other vector in P^k of length $|X|$ that makes angles θ_0 and θ_1 with V_0^k and V_1^k, respectively. (It is clear, by plane geometry, that there is no other such vector that is a linear combination of V_0^k and V_1^k; if a real vector is outside the plane of V_0^k and V_1^k, nowever, and θ is the angle between V_0^k and V_1^k, then we do not even have the angles with V_0^k and V_1^k summing to either θ or $2\pi - \theta$ as they must.)

Suppose $S_3(\omega') = S_3(\omega)$. As ω' must be a rearrangement of ω (by Fact 3 and B.6), we can take $Y \in R^k$ to b-represent ω'. By G.2,

$$X \cdot V_0^k = Y \cdot V_0^k$$
$$\& \quad X \cdot V_1^k = Y \cdot V_1^k$$
$$\& \quad |X| = |Y|.$$

Let θ_0' and θ_1' be the respective angles that Y makes with V_0^k and V_1^k. Since

$$X \cdot V_0^k = |X|\ |V_0^k| \cos \theta_0$$
$$\& \quad X \cdot V_1^k = |X|\ |V_1^k| \cos \theta_1$$

$$\& \quad Y \bullet V_0^{\ k} = |Y| \ |V_0^{\ k}| \cos \theta_0'$$

$$\& \quad Y \bullet V_1^{\ k} = |Y| \ |V_1^{\ k}| \cos \theta_1',$$

we can conclude that $\theta_0 = \theta_0'$ & $\theta_1 = \theta_1'$. As $|X| = |Y|$, then, we can conclude that $Y = X$, so that $\omega' = \omega$. □

Example. As $(1, 3, 5, 7) = 2V_1^{\ 4} - V_0^{\ 4}$, the binary string bababab is 3-local.

The obvious next questions is whether we have a converse of the last result; i.e., if ω is a 3-local binary string b-represented by $X_b \in R^{k_1}$ and a-represented by $X_a \in R^{k_2}$, need we have X_b a linear combination of $V_0^{\ k_1}$, $V_1^{\ k_1}$ or X_a a linear combination of $V_0^{\ k_2}$, $V_1^{\ k_2}$? (The answer is a trivial "yes" if $k_1 < 3$ or $k_2 < 3$, of course.) If this converse did hold, then we would have a complete and simple characterization of 3-localness; but the following result disappoints us:

Theorem G.14. Let $k_1 \geq 3$, $k_2 \geq 3$. Then we can find a 3-local rearrangement of $b^{k_1}a^{k_2}$ whose b-representation is linearly independent of $V_0^{\ k_1}$, $V_1^{\ k_1}$ and whose a-representation is linearly independent of $V_0^{\ k_2}$, $V_1^{\ k_2}$.

Proof. Let $\omega = b^{k_1-1} \ a^{k_2-1}$ ba, and let

$$X = (1, 2, \ldots, k_1-1, k_1+k_2-1),$$

$$Y = (k_1, k_1+1, \ldots, k_1+k_2-2, k_1+k_2).$$

Then X b-represents ω, and Y a-represents ω.

Suppose $X = t_0V_0^{\ k_1} + t_1V_1^{\ k_1}$. As $k_1 \geq 3$, we have at least the following:

$$t_0 + t_1 = 1$$

$$t_0 + 2t_1 = 2,$$

$$t_0 + k_1t_1 = k_1 + k_2-1.$$

From the first two equations, we get $t_1 = 1$ & $t_0 = 0$, so that the third equation gives $k_1 = k_1 + k_2 - 1$, contradicting $k_2 \geq 3$. Therefore, X is no linear combination of $V_0{}^{k_1}$ and $V_1{}^{k_1}$.

Next, suppose $Y = t_0 V_0{}^{k_2} + t_1 V_1{}^{k_2}$. As $k_2 \geq 3$, we have at least the following:

$$t_0 + t_1 = k_1,$$
$$t_0 + 2t_1 = k_1 + 1,$$
$$t_0 + k_2 t_1 = k_1 + k_2.$$

From the first two equations, we get $t_1 = 1$ & $t_0 = k_1 - 1$, so that the third equation gives $k_1 + k_2 - 1 = k_1 + k_2$, a contradiction. Therefore, Y is no linear combination of $V_0{}^{k_2}$ and $V_1{}^{k_2}$.

Now, let $S_3(\omega') = S_3(\omega)$. (Then ω' is some rearrangement of ω, by Fact 3 and B.6.) Let U b-represent ω'. Since $(S_3(\omega))([aba]) = k_2 - 1$, $(S_3(\omega'))([aba]) = k_2 - 1$. From this, we can infer that there is exactly one b mixed in among the a's of ω' and that it occurs just one position from the end of the a's. It follows that ω' is of one of two forms:

(i) $b^{\alpha} aba^{k_2-1} b^{\beta}$, where $\alpha + \beta = k_1 - 1$;

(ii) $b^{\alpha} a^{k_2-1} bab^{\beta}$, where $\alpha + \beta = k_1 - 1$.

First, suppose ω' is of form (i). A straightforward induction shows that $\Sigma u_i - \Sigma x_i = (\beta-1)(k_2-2)$. By G.2, we must have $\beta = 1$. But then $\Sigma u_i{}^2 - \Sigma x_i{}^2 = 4k_1 + 2k_2 - 2 \neq 0$, contradicting G.2. Therefore ω' cannot be of form (i).

Suppose ω' is of form (ii). A straightforward induction shows that $\Sigma u_i - \Sigma x_i = \beta k_2$. By G.2, we must have $\beta = 0$; i.e., we must have $\omega' = \acute{\omega}$. □

Now we turn to the following (already answered) question: For which

ℓ can we find ℓ many binary strings realizing the same 3-spectrum? I.e., how badly can strings fail to be 3-local? As there are strings that are not 3-local, D.2 shows that we can find arbitrarily many strings realizing the same 3-spectrum. We can make the question more difficult, however; Given ℓ, for which binary strings ω are there ℓ many rearrangements of ω all realizing the same 3-spectrum? In the present investigation, we actually look at a question that involves only "half of the 1-spectrum": Given ℓ, for which k can we find ℓ many strings, each containing exactly k many b's, that realize the same 3-spectrum? Along the way, a slightly less interesting question also arises: For which sets of differences $\{D_1, D_2, \ldots, D_m\} \subseteq Z^k$ do there exist vectors $X \in R^k$ such that

$$X,\ X + D_1,\ X + D_2,\ \ldots,\ X + D_m$$

b-represent strings realizing the same 3-spectrum? We do not arrive at complete answers to these questions, but we do give some interesting answers and discover some useful techniques along the way.

<u>Lemma G.15</u>. Let $X, Y \in R^k$, $t \in R$. If $\Sigma x_i = \Sigma y_i$ & $\Sigma i x_i = \Sigma i y_i$ & $\Sigma x_i^2 = \Sigma y_i^2$, then each of the following hold:

(i) $\Sigma t x_i = \Sigma t y_i$

& $\Sigma i(t x_i) = \Sigma i(t y_i)$

& $\Sigma (t x_i)^2 = \Sigma (t y_i)^2$;

(ii) $\Sigma (x_i + t) = \Sigma (y_i + t)$

& $\Sigma i(x_i + t) = \Sigma i(y_i + t)$

& $\Sigma (x_i + t)^2 = \Sigma (y_i + t)^2$;

(iii) $\Sigma (x_i + it) = \Sigma (y_i + it)$

& $\Sigma i(x_i + it) = \Sigma i(y_i + it)$

& $\Sigma (x_i + it)^2 = \Sigma (y_i + it)^2$.

Remark. It is only parts (i), (ii) that we actually use. Part (iii) is included because it fits in here.

Proof. Assume $\Sigma x_i = \Sigma y_i$ & $\Sigma i x_i = \Sigma i y_i$ & $\Sigma x_i^2 = \Sigma y_i^2$.

(i) $\Sigma t x_i = t \Sigma x_i = t \Sigma y_i = \Sigma t y_i$.

$\Sigma i(t x_i) = t \Sigma i x_i = t \Sigma i y_i = \Sigma i(t y_i)$.

$\Sigma (t x_i)^2 = t^2 \Sigma x_i^2 = t^2 \Sigma y_i^2 = \Sigma (t y_i)^2$.

(ii) $\Sigma (x_i + t) = \Sigma x_i + \Sigma t = \Sigma y_i + \Sigma t = \Sigma (y_i + t)$.

$\Sigma i(x_i + t) = \Sigma i x_i + \Sigma i t = \Sigma i y_i + \Sigma i t = \Sigma i(y_i + t)$.

$$\begin{aligned}\Sigma (x_i + t)^2 &= \Sigma x_i^2 + 2t \Sigma x_i + \Sigma t^2 \\ &= \Sigma y_i^2 + 2t \Sigma y_i + \Sigma t^2 = \Sigma (y_i + t)^2.\end{aligned}$$

(iii) $\Sigma (x_i + it) = \Sigma x_i + \Sigma i t = \Sigma y_i + \Sigma i t = \Sigma (y_i + it)$.

$\Sigma i(x_i + it) = \Sigma i x_i + \Sigma i^2 t = \Sigma i y_i + \Sigma i^2 t = \Sigma i(y_i + it)$.

$$\begin{aligned}\Sigma (x_i + it)^2 &= \Sigma x_i^2 + 2t \Sigma i x_i + \Sigma (it)^2 \\ &= \Sigma y_i^2 + 2t \Sigma i y_i + \Sigma (it)^2 \\ &= \Sigma (y_i + it)^2. \ \square\end{aligned}$$

Lemma G.16. Let $X, D \in R^k$

(i) $\Sigma x_i = \Sigma x_i + d_i \Leftrightarrow \Sigma d_i = 0$.

(ii) $\Sigma i x_i = \Sigma i(x_i + d_i) \Leftrightarrow \Sigma i d_i = 0$.

(iii) $\Sigma x_i^2 = \Sigma (x_i + d_i)^2 \Leftrightarrow D \cdot X = -\frac{1}{2} |D|^2$.

Proof. (i), (ii) are immediate by subtraction. To see (iii) just observe that

$$\begin{aligned}\Sigma x_i^2 = \Sigma (x_i + d_i)^2 &\Leftrightarrow |X|^2 = |X + D|^2 \\ &\Leftrightarrow |X|^2 = |X|^2 + 2D \cdot X + |D|^2 \\ &\Leftrightarrow D \cdot X = -\frac{1}{2} |D|^2. \ \square\end{aligned}$$

We can usefully restate this result as follows:

Corollary G.17. Let $X, D \in R^k$, and suppose $\Sigma d_i = \Sigma i d_i = 0$. Then

$$\Sigma x_i = \Sigma(x_i + d_i)$$

$$\& \quad \Sigma i x_i = \Sigma i(x_i + d_i)$$

$$\& \quad \Sigma {x_i}^2 = \Sigma(x_i + d_i)^2$$

$$\Leftrightarrow \quad D \bullet X = -\frac{1}{2}|D|^2.$$

This presents us with an opportunity to find longer strings realizing the same 3-spectrum. We just find some rational $D \neq 0$ satisfying $\Sigma d_i = \Sigma i d_i = 0$ and then look for a solution to $D \bullet X = -\frac{1}{2}|D|^2$ in $Q^k \cap M^k$. Then we use G.15 (i), (ii) and G.2 to get our strings.

Examples. Let $k = 4$. Then we need $D \in Q^4$ such that

$$d_1 + d_2 + d_3 + d_4 = 0$$

$$\& \quad d_1 + 2d_2 + 3d_3 + 4d_4 = 0;$$

i.e., we need to have

$$d_1 = d_3 + 2d_4$$

$$\& \quad d_2 = -(2d_3 + 3d_4).$$

Also, we need $X \in Q^4$ such that

$$d_1x_1 + d_2x_2 + d_3x_3 + d_4x_4 = -\frac{1}{2}|D|^2$$

$$\& \quad x_1 < x_- < x_3 < x_4$$

$$\& \quad x_1 + d_1 < x_2 + d_2 < x_3 + d_3 < x_4 + d_4.$$

The inequalities simplify to

$$x_2 - x_1 > \max(\{d_1 - d_2, 0\})$$

$$\& \quad x_3 - x_2 > \max(\{d_2 - d_3, 0\})$$

$$\& \quad x_4 - x_3 > \max(\{d_3 - d_4, 0\}).$$

So let us try $d_3 = d_4 = 1$. Then $D = (3, -5, 1, 1)$, and we want

$$3x_1 - 5x_2 + x_3 + x_4 = -18$$

& $x_2 - x_1 > 9$

& $x_3 - x_2 > 0$

& $x_4 - x_3 > 0$.

If we now try $x_1 = 1$ and $x_2 = 11$, we see that some acceptable solutions are

(1, 11, 12, 22),

(1, 11, 13, 21),

(1, 11, 14, 20),

(1, 11, 16, 18).

The third of these possibilities, for example, tells us that binary strings of the same length b-represented by (1, 11, 14, 20) and (4, 6, 15, 21) realize the same 3-spectrum; e.g.,

S_3(baaaaaaaaabaabaaaaaba) = S_3(aaababaaaaaaaabaaaaab).

Observe that our equations are also satisfied by (0, 10, 15.5, 16.5), which gives us the pair (0, 10, 15.5, 16.5) and (3, 5, 16.5, 17.5). These real vectors do not b-represent any strings at all, but we can use G.15 (i), (ii) to conclude that

2 (0, 10, 15.5, 16.5) + (1, 1, 1, 1) = (1, 21, 32, 34),

2 (3, 5, 16.5, 17.5) + (1, 1, 1, 1) = (7, 11, 34, 36)

b-represent strings realizing the same 3-spectrum.

Another particularly short example can be found with D = (-1, 1, 1, -1): the pair (2, 3, 6, 9) and (1, 4, 7, 8). This gives

S_3(abbaabaab) = S_3(baabaabba).

In a similar way, we can try taking two differences, D_1 and D_2, and finding X such that X, X + D_1, and X + D_2 are b-representations for three binary strings realizing the same 3-spectrum.

Example. If we take $D_1 = (3, -5, 1, 1)$ and $D_2 = (10, -17, 4, 3)$, then we get linear equations and inequalities that are satisfied by (0, 154, 155, 597), for example. By G.15 (i) and G.2, then, the binary strings of length 602 b-represented by (2, 156, 157, 599), (5, 151, 158, 600), and (12, 139, 161, 602) all realize the same 3-spectrum.

As the method used in this last example adds nothing significant to our method involving just one difference, we do not incorporate it into any of our formal results. Using the method, however, we have learned something new:

Corollary G.18. We can have three binary strings, each containing only four b's, realize the same 3-spectrum.

Remark. Since baaabba is the shortest string we know of that fails to be 3-local, the three strings given to us by the method of D.2 each contain six b's. We have shown, then, that not so many b's are necessary. In the process, however, we have sacrificed string length; the strings obtained by the earlier method were of length fourteen, while the ones we happened to find in the example above are of length 602.

If we consider the method used in the examples above, we may be tempted to guess that we can find no more than five binary strings, each containing exactly four b's, that realize the same 3-spectrum; for that number would correspond to four equations by the method used in the examples. To get such a result, we would need the following (false, we shall see) conjecture:

> If $X_0, X_1, \ldots, X_\ell$ are distinct members of R^k that b-represent binary strings that all realize the same 3-spectrum, then

$X_1 - X_0, X_2 - X_0, \ldots, X_\ell - X_0$ are linearly independent.

If we had this conjecture as a result, we would be able to conclude that, if there are k many b's contained in a binary string, then there are at most k many other binary strings realizing the same 3-spectrum as that string. (We have seen (G.5 and G.12) that, for $k < 4$, this proposed maximum is never realized.)

Let us turn to the conjecture. We can indeed get the result for $\ell = 2$.

Lemma G.19. Let $D_1, D_2 \in R^k$, $D_1 \neq 0$, $D_2 \neq 0$, $D_1 \neq D_2$, $t \in R$. If $D_2 = tD_1$, then there is no $X \in R^k$ such that $|X|^2 = |X + D_1|^2 = |X + D_2|^2$.

Proof. Suppose $D_2 = tD_1$ and $|X|^2 = |X + D_1|^2 = |X + D_2|^2$. By G.16 (iii),

$$D_1 \bullet X = -\tfrac{1}{2}|D_1|^2$$

$$\& \qquad D_2 \bullet X = -\tfrac{1}{2}|D_2|^2.$$

As $D_2 = tD_1$, $(tD_1) \bullet X = -\frac{1}{2}|tD_1|^2$, so that $D_1 \bullet X = t(-\frac{1}{2}|D_1|^2)$. As $D_1 \bullet X = -\frac{1}{2}|D_1|^2$, we must have $t = 1$; but that contradicts $D_1 \neq D_2$. □

Corollary G.20. If X_0, X_1, X_2 are distinct members of R^k that b-represent binary strings realizing the same 3-spectrum, then $X_1 - X_0$ and $X_2 - X_0$ are linearly independent.

Proof. Immediate, by G.2 and G.19 (with X_0, $X_1 - X_0$, $X_2 - X_0$ playing the roles of X, D_1, D_2, respectively, in the latter). □

Finally, now, we prove a theorem that shows that, Corollary G.20 notwithstanding, the above conjecture fails in a spectacular manner.

Theorem G.21. Assume $\ell \geq 0$. Then we can find ℓ many binary strings, each containing exactly four b's, that realize the same 3-spectrum.

Proof. We show how to find distinct

$$X_1, X_2, \ldots, X_\ell \in Z^4 \cap M^4 \cap P^4$$

such that, for i, j $\in \{1, 2, \ldots, \ell\}$,

$$(X_i - X_j) \cdot V_0^4 = 0 \ \& \ (X_i - X_j) \cdot V_1^4 = 0 \ \& \ |X_i| = |X_j|.$$

If m is the greatest coordinate of any of $X_1, X_2, \ldots, X_\ell$, then G.2 guarantees that these real vectors b-represent ℓ many strings of length m, all realizing the same 3-spectrum.

Take $D_1 = (-1, 1, 1, -1)$, $D_2 = (-1, 3, -3, 1)$. Then $D_1 \cdot D_2 = D_1 \cdot V_0^4 = D_1 \cdot V_1^4 = D_2 \cdot V_0^4 = D_2 \cdot V_1^4 = 0$. Let $W = 3V_1^4 = (3, 6, 9, 12)$, and let

$$P = \{t_1 D_1 + t_2 D_2 \mid t_1, t_2 \in R\},$$
$$P' = \{W + X \mid X \in P\}.$$

If $t_1 D_1 + t_2 D_2 \in P$, then

$$(t_1 D_1 + t_2 D_2) \cdot V_0^4 = t_1 (D_1 \cdot V_0^4) + t_2 (D_2 \cdot V_0^4) = 0$$
$$\& \quad (t_1 D_1 + t_2 D_2) \cdot V_1^4 = t_1 (D_1 \cdot V_1^4) + t_2 (D_2 \cdot V_1^4) = 0.$$

Let $C' = \{X \in P' \mid |X - W| = 2\}$.

Now, look at any $X = W + t_1 D_1 + t_2 D_2 \in C'$.

We have

$$(X - W) \cdot W = (t_1 D_1 + t_2 D_2) \cdot 3V_1^4$$
$$= 3t_1 (D_1 \cdot V_1^4) + 3t_2 (D_2 \cdot V_1^4) = 0.$$

By elementry eucledean geometry, then X - W and W are "perpendicular"; i.e., $|X|^2 = |W|^2 + |X - W|^2$. Therefore, $|X| = \sqrt{|W|^2 + |X - W|^2} = \sqrt{270 + 4} = \sqrt{274}$.

Now, observe that $W + D_1 = (2, 7, 10, 11) \in C' \cap M^4$. By continuity, then, $C' \cap M^4$ has non-empty interior (in the relativized topology of R^4 to C').

We claim that the rational points in C' (that is, $Q^4 \cap C'$) are dense in C' (in the relativized topology). To show this, it suffices to show that the rational points in $C = \{X \in P \mid |X| = 2\}$ are dense in C (in the relativized topoloty); for $X \in C'$ iff $X - W \in C$. In other words, it suffices to show that, for any $t \in R$ and any $\varepsilon > 0$, we can find rational t_1, t_2 such that $|t - \frac{t_1}{t_2}| < \varepsilon$ and $|t_1 D_1 + t_2 D_2| = 2$. To do this, we adapt the method used to get pythagorean triples in number theory [2]:

Let t, ε be given.

$$|t_1 D_1 + t_2 D_2| = 2 \Leftrightarrow |t_1 D_1 + t_2 D_2|^2 = 4$$
$$\Leftrightarrow |D_1|^2 t_1^2 + 2D_1 \cdot D_2 t_1 t_2 + |D_2|^2 t_2^2 = 4$$
$$\Leftrightarrow 4t_1^2 + 2 \cdot 0 \cdot t_1 t_2 + 20t_2^2 = 4$$
$$\Leftrightarrow t_1^2 + 5t_2^2 = 1.$$

Therefore, we must solve

$$a^2 + 5b^2 = c^2, \ |t - \frac{a}{b}| < \varepsilon$$

in integers. We can do this as follows: Find integers r, s such that $|t - \frac{r^2 - 5s^2}{2rs}| < \varepsilon$, and let $a = r^2 - 5s^2$, $b = 2rs$, $c = r^2 + 5s^2$.

As $C' \cap M^4$ has non-empty interior and the rational points are dense in C', $C' \cap M^4$ contains infinitely many rational points. Let $Y_1, Y_2, \ldots, Y_\ell \in C' \cap M^4 \cap Q^4$. As $Y_1, Y_2, \ldots, Y_\ell \in P'$, their differences are all in P. Let c be the (positive) least common denominator of these points. Then $c\,Y_1, c\,Y_2, \ldots, c\,Y_\ell \in Z^4 \cap M^4$ all have the same length (namely $c\sqrt{274}$) and have all their pairwise differences in P. Let d be the least coordinate of any $c\,Y_i$, and let $D = (1 - d, 1 - d, \ldots, 1 - d)$. Let $X_i = c\,Y_i + D$ for $i = 1, 2, \ldots, \ell$. By G.15 (ii), this gives the X_i as desired. □

<u>Example</u>. (In this example, we refer freely to the above proof.) Suppose we want five binary strings, each containing exactly four b's, that realize the same 3-spectrum. It appears that, if $t_1 > t_2$ and $W + t_1 D_1 +$

$t_2 D_2 \in C'$, then $W + t_1 D_1 + t_2 D_2$ is close enough to $W + D_1$ so that it lies in M^4. (Some points in C' with $t_1 \le t_2$ may also be close enough, but we need only a sufficient condition.) We get such points by taking $s = 1$ & $r \ge 4$. (Again, we care only about getting a sufficient condition.) If we take $r = 4$ & $s = 1$, we get

$$t_1 = \frac{r^2 - 5s^2}{r^2 + 5s^2} = \frac{11}{21}$$

$$\& \quad t_2 = \frac{2rs}{r^2 + 5s^2} = \frac{8}{21}$$

If we take $r = 5$ & $s = 1$, we get

$$t_1 = \frac{2}{3}$$

$$\& \quad t_2 = \frac{1}{3}.$$

If we take $r = 6$ & $s = 1$, we get

$$t_1 = \frac{31}{41}$$

$$\& \quad t_2 = \frac{12}{41}.$$

If we take $r = 7$ & $s = 1$, we get

$$t_1 = \frac{22}{27}$$

$$\& \quad t_2 = \frac{7}{27}.$$

If we take $r = 8$ & $s = 1$, we get

$$t_1 = \frac{59}{69}$$

$$\& \quad t_2 = \frac{16}{69}.$$

This gives the following five points in $C' \cap M^4$.

$$\begin{aligned} Y_1 &= W + \frac{11}{21} D_1 + \frac{8}{21} D_2 \\ &= \frac{1}{21} ((63, 126, 189, 252) + (-11, 11, 11, -11) + (-8, 24, -24, 8)) \\ &= \frac{1}{21} (44, 161, 176, 249), \end{aligned}$$

$$Y = W + \frac{2}{3} D_1 + \frac{1}{3} D_2$$
$$= \frac{1}{3} ((9, 18, 27, 36) + (-2, 2, 2, -2) + (-1, 3, -3, 1))$$
$$= \frac{1}{3} (6, 23, 26, 35),$$

$$Y_3 = W + \frac{31}{41} D_1 + \frac{12}{41} D_2$$
$$= \frac{1}{41} ((123, 246, 369, 492) + (-31, 31, 31, -31) + (-12, 36, -36, 12))$$
$$= \frac{1}{41} (80, 313, 364, 473),$$

$$Y_4 = W + \frac{22}{27} D_1 + \frac{7}{27} D_2$$
$$= \frac{1}{27} ((81, 162, 243, 324) + (-22, 22, 22, -22) + (-7, 21, -21, 7))$$
$$= \frac{1}{27} (52, 205, 244, 309),$$

$$Y_5 = W + \frac{59}{69} D_1 + \frac{16}{69} D_2$$
$$= \frac{1}{69} ((207, 414, 621, 828) + (-59, 59, 59, -59) + (-16, 48, -48, 16))$$
$$= \frac{1}{69} (132, 521, 632, 785).$$

The least common denominator of these points is $c = 3^3 \cdot 7 \cdot 23 \cdot 41$, so

$$c\ Y_1 = 9 \cdot 23 \cdot 41\ (44, 161, 176, 249)$$
$$= (373428, 1366407, 1493712, 2113263),$$

$$c\ Y_2 = 9 \cdot 7 \cdot 23 \cdot 41\ (6, 23, 26, 35)$$
$$= (356354, 1366407, 1544634, 2079315),$$

$$c\ Y_3 = 27 \cdot 7 \cdot 23\ (80, 313, 364, 473)$$
$$= (347760, 1360611, 1582308, 2056131),$$

$$c\ Y_4 = 7 \cdot 23 \cdot 41\ (52, 205, 244, 309)$$
$$= (343252, 1353205, 1610644, 2039709),$$

$$c\ Y_5 = 9 \cdot 7 \cdot 41\ (132, 521, 632, 785)$$
$$= (340956, 1345743, 1632456, 2027655).$$

Finally, then, let

$$X_1 = c\ Y_1 - (340955, 340955, 340955, 340955)$$
$$= (32473, 1025452, 1152757, 1772308),$$

$$X_2 = c\,Y_2 - (340955,\ 340955,\ 340955,\ 340955)$$
$$= (15499,\ 1025452,\ 1203679,\ 1738360),$$
$$X_3 = c\,Y_3 - (340955,\ 340955,\ 340955,\ 340955)$$
$$= (6805,\ 1019656,\ 1241353,\ 1715176),$$
$$X_4 = c\,Y_4 - (340955,\ 340955,\ 340955,\ 340955)$$
$$= (2297,\ 1012250,\ 1269689,\ 1698754),$$
$$X_5 = c\,Y_5 - (340955,\ 340955,\ 340955,\ 340955)$$
$$= (1,\ 1004788,\ 1291501,\ 1686700).$$

These five real vectors b-represent five binary strings (of length 1772308) that realize the same 3-spectrum.

H. 4-events

Now we turn to the even harder task of finding and studying strings which are not 4-local (if such strings actually exist). We do not get as many results as in the previous section, but we do demonstrate a new technique and see the inspiration for some (as yet unsettled) conjectures.

The proof of our first result is actually a failing attempt to find a string of length eleven that is not 4-local. It illustrates a new technique for limiting such a search.

Theorem H.1. $G(11) \leq 4$.

Proof. Suppose $G(11) \nleq 4$. By E.7, $G(11) > 4$; so, by C.20, there are binary strings ω_1 and ω_2 of length eleven such that $\omega_1(0) = b \neq \omega_2(0)$ & $S_4(\omega_1) = S_4(\omega_2)$.

The number of 4-reports on a string of length eleven is

$$\binom{11}{4} = \frac{11!}{4!\ 7!} = 330.$$

Now, the number of these that start at the beginning of the string is

$$\binom{10}{3} = \frac{10!}{3!\ 7!}$$

Similarly, the respective numbers of these that start at the second, third, fourth, fifth, sixth, seventh, and eighth positions are 84, 56, 35, 20, 10, 4, and 1. Since $120 + 84 + 56 + 35 + 20 + 10 + 4 + 1 = 330$, this accounts for all the 4-reports on a string of length eleven.

$$\text{Let} \qquad m = \sum_{\substack{\overline{\overline{\omega}}=4 \\ \omega(0)=b}} (S_4(\omega_1))\ ([\omega]) = \sum_{\substack{\overline{\overline{\omega}}=4 \\ \omega(0)=b}} (S_4(\omega_2))\ ([\omega]).$$

As $\omega_1(0) = b$, $m - 120$ is the sum of the members of some subset of $\{84, 56, 35, 20, 10, 4, 1\}$; and, as $\omega_2(0) \neq b$, m is the sum of the members of some subset of $\{84, 56, 35, 20, 10, 4, 1\}$, too. For example we might have

$$m = 150 = 120 + 20 + 10 = 84 + 56 + 10$$

$$\text{or} \quad m = 140 = 120 + 20 = 84 + 56.$$

(In the former case, the initial length eight substrings of ω_1 and ω_2, respectively, are baaabbaa and abbaabaa; in the latter case, they are baaabaaa and abbaaaaa. We are about to see that neither these cases nor any others are real possibilities; at this point in the argument, however, they are among the "candidates.") In the latter case, the subsets of $\{84, 56, 35, 20, 10, 4, 1\}$ summing to m - 120 and m are disjoint (meaning $\omega_2(i) = \omega_1(i) \Rightarrow \omega_1(i) = a$, for $i \in 8$); let us call such cases "primitive." Observe that the former case can be obtained from the latter case by adding the unused number ten to both sums. Clearly, any case can be obtained from some unique primitive case by adding some unused members of $\{84, 56, 35, 20, 10, 4, 1\}$. An exhaustive search reveals only the following possible primitive cases:

(i) $m = 120 = 84 + 35 + 1$;

(ii) $m = 121 = 120 + 1 = 56 + 35 + 20 + 10$;

(iii) $m = 140 = 120 + 20 = 84 + 56$;

(iv) $m = 155 = 120 + 35 = 84 + 56 + 10 + 4 + 1$;

(v) $m = 160 = 120 + 35 + 4 + 1 = 84 + 56 + 20$;

(vi) $m = 165 = 120 + 35 + 10 = 84 + 56 + 20 + 4 + 1$.

Let X and Y b-represent ω_1 and ω_2, respectively. By Fact 3 and G.2, we must have $\bar{\bar{X}} = \bar{\bar{Y}}$ & $\Sigma x_i^2 = \Sigma y_i^2$. Below, we rule this our in each case, however, arriving at a contradiction.

(i) Assume the primitive case from which ω_1 and ω_2 are derived is given by (i) above. Then we must have

$$1^2 + \ldots = 2^2 + 4^2 + 8^2 + \ldots,$$

where the rest of each sum depends on the end of the corresponding string and must be one of the following:

$0,$
$9^2,$
$10^2,$
$11^2,$
$9^2 + 10^2,$
$9^2 + 11^2,$
$10^2 + 11^2,$
$9^2 + 10^2 + 11^2.$

Hence, $(2^2 + 4^2 + 8^2) - 1^2 = 83$ must be one of the pairwise differences among these eight numbers; i.e., eighty-three must be among the following fourteen numbers:

$0,$
$9^2 = 81,$
$10^2 = 100,$
$11^2 = 121,$
$9^2 + 10^2 = 181,$
$9^2 + 11^2 = 202,$
$10^2 + 11^2 = 221,$
$9^2 + 10^2 + 11^2 = 302,$
$10^2 - 9^2 = 19,$
$11^2 - 9^2 = 40,$
$10^2 + 11^2 - 9^2 = 140,$
$11^2 - 10^2 = 21,$
$9^2 + 11^2 - 10^2 = 102,$
$9^2 + 10^2 - 11^2 = 60.$

As it is not, this case is ruled out.

(ii) Assume the primitive case from which ω_1 and ω_2 are derived is given by (ii) above. Now, $(3^2 + 4^2 + 5^2 + 6^2) - (1^2 + 8^2) = 21$ is among the above listed differences $(21 = 11^2 - 10^2 = (9^2 + 11^2) - (9^2 + 10^2))$;

but then the final length-three substrings of ω_1 and ω_2 must be rearrangements, so that ω_2 must contain two more b's than ω_1 does, a contradiction. Therefore, this case is ruled out, too.

(iii) Assume the primitive case from which ω_1 and ω_2 are derived is given by (iii) above. As $(1^2 + 5^2) - (2^2 + 3^2) = 13$ is not among the above listed differences, this case is ruled out.

(iv) Assume the primitive case from which ω_1 and ω_2 are derived is given by (iv) above. As $(2^2 + 3^2 + 6^2 + 7^2 + 8^2) - (1^2 + 4^2) = 145$ is not among the above listed differences, this case is ruled out.

(v) Assume the primitive case from which ω_1 and ω_2 are derived is given by (v) above. As $(1^2 + 4^2 + 7^2 + 8^2) - (2^2 + 3^3 + 5^2) = 92$ is not among the above listed differences, this case is ruled out.

(vi) Assume the primitive case from which ω_1 and ω_2 are derived is given by (vi) above. As $(2^2 + 3^2 + 5^2 + 7^2 + 8^2) - (1^2 + 4^2 + 6^2) = 98$ is not among the above listed differences, this case is ruled out. □

Corollary H.2. (i) $G(9) = 4$.

(ii) $G(10) = 4$.

(iii) $G(11) = 4$.

Proof. We observed in the remark on page 63 that $G(8) = 4$. By H.1, $G(11) \leq 4$; so these results follow by Fact 8. □

Corollary H.3. If there are strings that are not 4-local, then they are longer than eleven.

Proof. Immediate, as $G(11) = 4$. □

This last result is rather discouraging, and one might begin to suspect that every string is 4-local. Rather than blindly continue to look at binary strings of length twelve, then, let us seek a characterization of 4-localness that might allow us to prove the futility of our search if it is indeed futile.

Theorem H.4. Let ω_1 and ω_2 be binary strings realizing the same 3-spectrum, and let X and Y b-represent ω_1 and ω_2, respectively. Then $S_4(\omega_1) = S_4(\omega_2)$ iff

$$\Sigma i^2 x_i = \Sigma i^2 y_i$$

$$\& \quad \Sigma i x_i^2 = \Sigma i y_i^2$$

$$\& \quad \Sigma x_i^3 = \Sigma y_i^3.$$

Proof. By G.2, ω_1 and ω_2 are rearrangements with

$$\Sigma x_i = \Sigma y_i$$

$$\& \quad \Sigma i x_i = \Sigma i y_i$$

$$\& \quad \Sigma x_i^2 = \Sigma y i^2.$$

Let $\ell = \bar{\bar{\omega}}_1 = \bar{\bar{\omega}}_2$, and let $k = \bar{\bar{X}} = \bar{\bar{Y}}$. As $S_3(\omega_1) = S_3(\omega_2)$, $S_4(\omega_1) = S_4(\omega_2)$ iff $S_4(\omega_1)$ and $S_4(\omega_2)$ agree on all 4-events. The strings ω_1 and ω_2 are rearrangements, so $S_4(\omega_1)$ and $S_4(\omega_2)$ agree on [bbbb], [aaaa]. The only other 4-events are [abbb], [aabb], [aaab], [bbba], [bbaa], [baaa], [abaa], [aaba], [babb], [bbab], [baba], [abab], [abba], [baab].

For each i (i = 1, 2, ..., k) let f(i), f′(i), g(i), g′(i) be the respective cardinalities of

$$\{j < x_i - 1 \mid \omega_1(j) = b\},$$

$$\{j < x_i - 1 \mid \omega_1(j) = a\},$$

$$\{j > x_i - 1 \mid \omega_1(j) = b\},$$

$$\{j > x_i - 1 \mid \omega_1(j) = a\}.$$

Explicitly, then,

$$f(i) = i - 1,$$

$$f'(i) = x_i - i,$$

$g(i) = k - i,$

$g'(i) = (\ell - k) - (x_i - i).$

Now,

$$(S_4(\omega_1))\ ([abbb]) = \sum_{i=1}^{k} f'(i) \binom{g(i)}{2}$$

$$= \sum_{i=1}^{k} (x_i - i) \frac{(k-i)\ (k-i-1)}{2!}$$

$$= \frac{1}{2} \sum_{i=1}^{k} x_i\ (k^2 - k + (1-2k)i + i^2)$$

$$- \frac{1}{2} \sum_{i=1}^{k} i\ (k^2 - k + (1-2k)i + i^2)$$

$$= \frac{1}{2}\ (k^2 - k) \sum_{i=1}^{k} x_i + \frac{1}{2}\ (1-2k) \sum_{i=1}^{k} ix_i + \frac{1}{2} \sum_{i=1}^{k} i^2 x_i$$

$$- \frac{1}{2} \sum_{i=1}^{k} i\ (k^2 - k + (1-2k)i + i^2).$$

Similarly,

$$(S_4(\omega_2))\ ([abbb]) = \frac{1}{2}\ (k^2 - k)\ \Sigma y_i + \frac{1}{2}\ (1 - 2k)\ \Sigma i y_i$$

$$+ \frac{1}{2}\ \Sigma i^2 y_i - \frac{1}{2}\ \Sigma i(k^2 - k + (1-2k)i + i^2).$$

As $\Sigma x_i = \Sigma y_i$ & $\Sigma i x_i = \Sigma i y_i$ & $\Sigma x_i^2 \Rightarrow \Sigma y_i^2$, then,

$$(S_4(\omega_1))\ ([abbb]) = (S_4(\omega_2))\ ([abbb])$$

$$\Leftrightarrow \Sigma i^2 x_i = \Sigma i^2 y_i.$$

Continuing our calculations, we get

$$(S_4(\omega_1))\ ([aabb]) = \sum_{i=1}^{k} \binom{f'(i)}{2}\ g(i)$$

$$= \sum_{i=1}^{k} \frac{(x_i - i)(x_i - i - 1)}{2!}\ (k-i)$$

$$= \frac{1}{2}\sum_{i=1}^{k} k(x_i^2 - 2ix_i + i^2 - x_i + 1)$$

$$- \frac{1}{2}\sum_{i=1}^{k} i(x_i^2 - 2ix_i + i^2 - x_i + 1)$$

$$= \frac{1}{2}k\sum_{i=1}^{k} x_i^2 - k\sum_{i=1}^{k} ix_i + \frac{1}{2}k\sum_{i=1}^{k} i^2 - \frac{1}{2}k\sum_{i=1}^{k} x_i + \frac{1}{2}k^2$$

$$- \frac{1}{2}\sum_{i=1}^{k} ix_i^2 + \sum_{i=1}^{k} i^2 x_i - \frac{1}{2}\sum_{i=1}^{k} i^3 + \frac{1}{2}\sum_{i=1}^{k} ix_i - \frac{1}{2}\sum_{i=1}^{k} i,$$

and

$$(S_4(\omega_2))\ ([aabb]) =$$

$$\frac{1}{2}\,k\,\Sigma y_i^2 - k\,\Sigma iy_i + \frac{1}{2}k\Sigma i^2 - \frac{1}{2}\,k\,\Sigma y_i + \frac{1}{2}\,k^2$$

$$- \frac{1}{2}\,\Sigma iy_i^2 + \Sigma i^2 y_i - \frac{1}{2}\,\Sigma i^3 + \frac{1}{2}\,\Sigma iy_i - \frac{1}{2}\,\Sigma i.$$

Therefore,

$$(S_4(\omega_1))\ ([aabb]) = (S_4(\omega_2))\ ([aabb])$$

$$\Leftrightarrow \Sigma i^2 x_i - \frac{1}{2}\,\Sigma ix_i^2 = \Sigma i^2 y_i - \frac{1}{2}\,\Sigma iy_i^2.$$

Continuing,

$$(S_4(\omega_1))\ ([aaab]) = \sum_{i=1}^{k} \binom{f'(i)}{3}$$

$$= \sum_{i=1}^{k} \frac{(x_i - i)(x_i - i - 1)(x_i - i - 2)}{3!}$$

$$= \frac{1}{6}\sum_{i=1}^{k} x_i \, (x_i^2 - 2ix_i + i^2 - x_i + 1)$$

$$- \frac{1}{6}\sum_{i=1}^{k} i \, (x_i^2 - 2ix_i + i^2 - x_i + 1)$$

$$- \frac{1}{3}\sum_{i=1}^{k} (x_i^2 - 2ix_i + i^2 - x_i + 1)$$

$$= \frac{1}{6}\sum_{i=1}^{k} x_i^3 - \frac{1}{3}\sum_{i=1}^{k} ix_i^2 + \frac{1}{6}\sum_{i=1}^{k} i^2 x_i - \frac{1}{6}\sum_{i=1}^{k} x_i^2 + \frac{1}{6}\sum_{i=1}^{k} x_i$$

$$- \frac{1}{6}\sum_{i=1}^{k} ix_i^2 + \frac{1}{3}\sum_{i=1}^{k} i^2 x_i - \frac{1}{6}\sum_{i=1}^{k} i^3 + \frac{1}{6}\sum_{i=1}^{k} ix_i - \frac{1}{6}\sum_{i=1}^{k} i$$

$$- \frac{1}{3}\sum_{i=1}^{k} x_i^2 + \frac{2}{3}\sum_{i=1}^{k} ix_i - \frac{1}{3}\sum_{i=1}^{k} i^2 + \frac{1}{3}\sum_{i=1}^{k} x_i - \frac{1}{3}k,$$

and

$$(S_4(\omega_2))([aaab]) =$$

$$\frac{1}{6}\,\Sigma y_i^3 - \frac{1}{3}\,\Sigma iy_i^2 + \frac{1}{6}\,\Sigma i^2 y_i - \frac{1}{6}\,\Sigma y_i^2 + \frac{1}{6}\,\Sigma y_i$$

$$- \frac{1}{6}\,\Sigma iy_i^2 + \frac{1}{3}\,\Sigma i^2 y_i - \frac{1}{6}\,\Sigma i^3 + \frac{1}{6}\,\Sigma iy_i - \frac{1}{6}\Sigma i$$

$$- \frac{1}{3}\,\Sigma y_i^2 + \frac{2}{3}\,\Sigma iy_i - \frac{1}{3}\,\Sigma i^2 + \frac{1}{3}\,\Sigma y_i - \frac{1}{3}k.$$

Therefore,

$$(S_4(\omega_1))([aaab]) = (S_4(\omega_2))([aaab])$$

$$\Leftrightarrow \frac{1}{2}\,\Sigma i^2 x_i - \frac{1}{2}\,\Sigma ix_i^2 + \frac{1}{6}\,\Sigma x_i^3 = \frac{1}{2}\,\Sigma i^2 y_i - \frac{1}{2}\,\Sigma iy_i^2 + \frac{1}{6}\,\Sigma y_i^3.$$

Accumulating our findings, we get

$$S_4(\omega_1) = S_4(\omega_2) \Leftrightarrow$$

$$\Sigma i^2 x_i = \Sigma i^2 y_i$$

$$\& \ \Sigma i^2 x_i - \frac{1}{2} \Sigma i x_i^2 = \Sigma i^2 y_i - \frac{1}{2} \Sigma i y_i^2$$

$$\& \ \frac{1}{2} \Sigma i^2 x_i - \frac{1}{2} \Sigma i x_i^2 + \frac{1}{6} \Sigma x_i^3 = \frac{1}{2} \Sigma i^2 y_i - \frac{1}{2} \Sigma i y_i^2 + \frac{1}{6} \Sigma y_i^3$$

& $S_4(\omega_1)$, $S_4(\omega_2)$ agree on [bbba], [bbaa], [baaa], [abaa],

[aaba], [babb], [bbab], [baba], [abab], [abba], [baab].

Simplifying,

$$S_4(\omega_1) = S_4(\omega_2) \Leftrightarrow$$

$$\Sigma i^2 x_i = \Sigma i^2 y_i \ \& \ \Sigma i x_i^2 = \Sigma i y_i^2 \ \& \ \Sigma x_i^3 = \Sigma y_i^3$$

& $S_4(\omega_1)$, $S_4(\omega_2)$ agree on [bbba], [bbaa], [baaa], [abaa],

[aaba], [babb], [bbab], [baba], [abab], [abba], [baab].

It will suffice, then, to show that

$$\Sigma i^2 x_i = \Sigma i^2 y_i \ \& \ \Sigma i x_i^2 = \Sigma i y_i^2 \ \& \ \Sigma x_i^3 = \Sigma y_i^3$$

$\Rightarrow$ $S_4(\omega_1)$, $S_4(\omega_2)$ agree on [bbba], [bbaa], [baaa], [abaa],

[aaba], [babb], [bbab], [baba], [abab], [abba], [baab].

So assume $\Sigma i^2 x_i = \Sigma i^2 y_i$ & $\Sigma i x_i^2 = \Sigma i y_i^2$ & $\Sigma x_i^3 = \Sigma y_i^3$. As we also have

$$\Sigma x_i = \Sigma y_i \ \& \ \Sigma i x_i = \Sigma i y_i \ \& \ \Sigma x_i^2 = \Sigma y_i^2,$$

we know that, if P(i,x) is any polynomial in i and x having degree at most three, then $\sum_{i=1}^{k} P(i,x_i) = \sum_{i=1}^{k} P(i,y_i)$; we use this fact to show that $S_4(\omega_1)$, $S_4(\omega_2)$ agree on [bbba], [bbaa], [baaa], [abaa], [aaba], [babb], [bbab], [baba], [abab], [abba], [baab].

$$(S_4(\omega_1))\ ([bbba]) = \sum_{i=1}^{k} \binom{f(i)}{2} g'(i)$$

$$= \frac{1}{2} \sum_{i=1}^{k} (i-1)\ (i-2)\ [(\ell-k) - (x_i-i)]$$

$$= \frac{1}{2} \sum_{i=1}^{k} (i-1)\ (i-2)\ [(\ell-k) - (y_i-i)]$$

$$= (S_4(\omega_2))\ ([bbba]).$$

$$(S_4(\omega_1))\ ([bbaa]) = \sum_{i=1}^{k} f(i) \binom{g'(i)}{2}$$

$$= \frac{1}{2} \sum_{i=1}^{k} (i-1)\ [(\ell-k) - (x_i-i)]\ [(\ell-k) - (x_i-i) - 1]$$

$$= \frac{1}{2} \sum_{i=1}^{k} (i-1)\ [(\ell-k) - (y_i-i)]\ [(\ell-k) - (y_i-i) - 1]$$

$$= (S_4(\omega_2))\ ([bbaa]).$$

$$(S_4(\omega_1))\ ([baaa]) = \sum_{i=1}^{k} \binom{g'(i)}{3}$$

$$= \frac{1}{6} \sum_{i=1}^{k} [(\ell-k) - (x_i-i)]\ [(\ell-k) - (x_i-i) - 1]\ [(\ell-k) - (x_i-i) - 2]$$

$$= \frac{1}{6} \sum_{i=1}^{k} [(\ell-k) - (y_i-i)]\ [(\ell-k) - (y_i-i) - 1]\ [(\ell-k) - (y_i-i) - 2]$$

$$= (S_4(\omega_2))\ ([baaa]).$$

$$(S_4(\omega_1))\ ([abaa]) = \sum_{i=1}^{k} f'(i) \binom{g'(i)}{2}$$

$$= \frac{1}{2}\sum_{i=1}^{k} (x_i - i)\ [(\ell-k) - (x_i - i)]\ [(\ell-k) - (x_i - i) - 1]$$

$$= \frac{1}{2}\sum_{i=1}^{k} (y_i - i)\ [(\ell-k) - (y_i - i)]\ [(\ell-k) - (y_i - i) - 1]$$

$$= (S_4(\omega_2))\ ([abaa]).$$

$$(S_4(\omega_1))\ ([aaba]) = \Sigma \binom{f'(i)}{2} g'(i)$$

$$= \frac{1}{2}\sum_{i=1}^{k} (x_i - i)\ (x_i - i - 1)\ [(\ell-k) - (x_i - i)]$$

$$= \frac{1}{2}\sum_{i=1}^{k} (y_i - i)\ (y_i - i - 1)\ [(\ell-k) - (y_i - i)]$$

$$= (S_4(\omega_2))\ ([aaba]).$$

$$(S_4(\omega_1))\ ([babb]) = \sum_{i=1}^{k}\sum_{j=i}^{k} [f'(j) - f'(i)]\ g(j)$$

$$= \sum_{i=1}^{k}\sum_{j=i}^{k} [(x_j - j) - (x_i - i)]\ (k-j)$$

$$= k\sum_{i=1}^{k}\sum_{j=i}^{k} (x_j - j) - \sum_{i=1}^{k}\sum_{j=i}^{k} (x_j - j)\ j$$

$$- k\sum_{i=1}^{k} (x_i - i)\ (k-i+1) + \sum_{i=1}^{k} (x_i - i) \left(\sum_{j=i}^{k} j\right)$$

$$= k\sum_{i=1}^{k} i(x_i-i)$$

$$-\sum_{i=1}^{k} i(x_i-i)\ i$$

$$- k\sum_{i=1}^{k} (x_i-i)\ (k-i+1)$$

$$+\sum_{i=1}^{k} (x_i-i)\ \frac{(k-i+1)\ (i+k)}{2}$$

$$= k\sum_{i=1}^{k} i(y_i-i)$$

$$-\sum_{i=1}^{k} i(y_i-i)\ i$$

$$- k\sum_{i=1}^{k} (y_i-i)\ (k-i+1)$$

$$+\sum_{i=1}^{k} (y_i-i)\ \frac{(k-i+1)\ (i+k)}{2}$$

$$= (S_4(\omega_2))\ ([babb]).$$

$$(S_4(\omega_1))\ ([bbab]) = \sum_{i=1}^{k} f(i) \left(\sum_{j=i}^{k} [f'(j) - f'(i)]\right)$$

$$= \sum_{i=1}^{k} (i-1) \left(\sum_{j=i}^{k} [(x_j-j) - (x_i-i)]\right)$$

$$= \sum_{i=1}^{k} (i-1) \left(\sum_{j=i}^{k} (x_j-j)\right) - \sum_{i=1}^{k} (i-1)\ (x_i-i)\ (k-i+1)$$

$$= \sum_{i=1}^{k} \frac{i(i-1)}{2}\ (x_i-i) - \sum_{i=1}^{k} (i-1)\ (x_i-i)\ (k-i+1)$$

$$= \sum_{i=1}^{k} \frac{i(i-1)}{2}\ (y_i-i) - \sum_{i=1}^{k} (i-1)\ (y_i-i)\ (k-i+1)$$

$$= (S_4(\omega_2))\ ([bbab]).$$

$$(S_4(\omega_1))\ ([baba]) = \sum_{i=1}^{k} \sum_{j=i}^{k} [f'(j) - f'(i)]\ g'(j)$$

$$= \sum_{i=1}^{k} \sum_{j=i}^{k} [(x_j - j) - (x_i - i)]\ [(\ell - k) - (x_j - j)]$$

$$= (\ell - k) \sum_{i=1}^{k} \sum_{j=i}^{k} (x_j - j)$$

$$- \sum_{i=1}^{k} \sum_{j=i}^{k} (x_j - j)^2$$

$$- (\ell - k) \sum_{i=1}^{k} (x_i - i)\ (k - i + 1)$$

$$+ \sum_{i=1}^{k} (x_i - i) \left(\sum_{j=i}^{k} (x_j - j) \right)$$

$$= (\ell - k) \sum_{i=1}^{k} i\ (x_i - i)$$

$$- \sum_{i=1}^{k} i (x_i - i)^2$$

$$- (\ell - k) \sum_{i=1}^{k} (x_i - i)\ (k - i + 1)$$

$$+ \frac{1}{2} \left[\left(\sum_{i=1}^{k} (x_i - i) \right)^2 + \sum_{i=1}^{k} (x_i - i)^2 \right]$$

$$= (\ell - k) \sum_{i=1}^{k} i (y_i - i)$$

$$- \sum_{i=1}^{k} i (y_i - i)^2$$

$$- (\ell - k) \sum_{i=1}^{k} (y_i - i)\ (k - i + 1)$$

$$+ \frac{1}{2}\left[\left(\sum_{i=1}^{k}(y_i - i)\right)^2 + \sum_{i=1}^{k}(y_i - i)^2\right]$$

$$= \quad (S_4(\omega_2))\ ([baba]).$$

$$(S_4(\omega_1))\ ([abab]) = \sum_{i=1}^{k} f'(i) \left(\sum_{j=i}^{k} [f'(j) - f'(i)]\right)$$

$$= \quad \sum_{i=1}^{k}(x_i - i) \left(\sum_{j=i}^{k} [(x_j - j) - (x_i - i)]\right)$$

$$= \quad \sum_{i=1}^{k}(x_i - i) \left(\sum_{j=i}^{k}(x_j - j)\right)$$

$$- \sum_{i=1}^{k}(x_i - i)^2\ (k-i+1)$$

$$= \quad \frac{1}{2}\left[\left(\sum_{i=1}^{k}(x_i - i)\right)^2 + \sum_{i=1}^{k}(x_i - i)^2\right] - \sum_{i=1}^{k}(x_i - i)^2\ (k-i+1)$$

$$= \quad \frac{1}{2}\left[\left(\sum_{i=1}^{k}(y_i - i)\right)^2 + \sum_{i=1}^{k}(y_i - i)^2\right] - \sum_{i=1}^{k}(y_i - i)^2\ (k-i+1)$$

$$= \quad (S_4(\omega_2))\ ([abab]).$$

$$(S_4(\omega_1))\ ([abba]) = \sum_{i=1}^{k} f'(i) \left(\sum_{j=i+1}^{k} g'(j)\right)$$

$$= \quad \sum_{i=1}^{k}(x_i - i) \left(\sum_{j=i+1}^{k} [(\ell - k) - (x_j - j)]\right)$$

$$= \quad (\ell - k) \sum_{i=1}^{k}(x_i - i)\ (k-i)$$

$$- \sum_{i=1}^{k}(x_i - i) \left(\sum_{j=i+1}^{k}(x_j - j)\right)$$

$$= \quad (\ell - k) \sum_{i=1}^{k}(x_i - i)\ (k-i)$$

$$- \frac{1}{2}\left[\left(\sum_{i=1}^{k} (x_i-i)\right)^2 - \sum_{i=1}^{k}(x_i-i)^2\right]$$

$$= \quad (\ell-k) \sum_{i=1}^{k}(y_i-i)\ (k-i)$$

$$- \frac{1}{2}\left[\left(\sum_{i=1}^{k}(y_i-i)\right)^2 - \sum_{i=1}^{k}(y_i-i)^2\right]$$

$$= \quad (S_4(\omega_2))\ ([abba]).$$

(Observe that here we needed only

$$\Sigma x_i = \Sigma y_i \ \&\ \Sigma i x_i = \Sigma i y_i \ \&\ \Sigma x_i^2 = \Sigma y_i^2.$$

This gives us Corollary H.5.)

$$(S_4(\omega_1))\ ([baab]) = \sum_{i=1}^{k} \sum_{j=i}^{k} \binom{f'(j) - f'(i)}{2}$$

$$= \quad \frac{1}{2}\sum_{i=1}^{k}\sum_{j=i}^{k} [(x_j-j) - (x_i-i)]\ [(x_j-j) - (x_i-i) - 1]$$

$$= \quad \frac{1}{2}\sum_{i=1}^{k}\sum_{j=i}^{k}(x_j-j)^2$$

$$- \sum_{i=1}^{k}(x_i-i)\left(\sum_{j=i}^{k}(x_j-j)\right)$$

$$+ \frac{1}{2}\sum_{i=1}^{k}(x_i-i)^2\ (k-i+1)$$

$$- \sum_{i=1}^{k} \sum_{j=i}^{k}(x_j-j) + \sum_{i=1}^{k}(x_i-i)\ (k-i+1)$$

$$= \quad \frac{1}{2}\sum_{i=1}^{k} i(x_i-i)^2$$

$$- \frac{1}{2}\left[\left(\sum_{i=1}^{k}(x_i-i)\right)^2 + \sum_{i=1}^{k}(x_i-i)^2\right]$$

$$+\frac{1}{2}\sum_{i=1}^{k}(x_i-i)^2\ (k-i+1)$$

$$-\sum_{i=1}^{k} i(x_i-i) + \sum_{i=1}^{k}(x_i-i)\ (k-i+1)$$

$$= \frac{1}{2}\sum_{i=1}^{k} i(y_i-i)^2$$

$$-\frac{1}{2}\left[\left(\sum_{i=1}^{k}(y_i-i)\right)^2 + \sum_{i=1}^{k}(y_i-i)^2\right]$$

$$+\frac{1}{2}\sum_{i=1}^{k}(y_i-i)^2\ (k-i+1)$$

$$-\sum_{i=1}^{k} i(y_i-i) + \sum_{i=1}^{k}(y_i-i)\ (k-i+1)$$

$$= (S_4(\omega_2))\ ([baab]).\ \square$$

Remark. The simplicity of this result (H.4) suggests that there might be a less messy proof, but attempts to find easy proofs of this and the earlier results F.1 and G.1 have not yet succeeded.

Before giving our main corollary, we record two incidental corollaries to the proof of H.4. These show that there are 4-events other than [bbbb], and [aaaa] whose occurrences in a binary string are determined by the 3-spectrum realized by the string.

Corollary H.5. If ω_1 and ω_2 are binary strings realizing the same 3-spectrum, then $(S_4(\omega_1))\ ([abba]) = (S_4(\omega_2))\ ([abba])$.

Proof. Let ω_1 and ω_2 be binary strings realizing the same 3-spectrum and b-represented by X and Y, respectively. By G.2, ω_1 and ω_2 are rearrangements with $\Sigma x_i = \Sigma y_i$ & $\Sigma ix_i = \Sigma iy_i$ & $\Sigma x_i^2 = \Sigma y_i^2$. Let $\ell = \bar{\bar{\omega}}_1 = \bar{\bar{\omega}}_2$, and let $k = \bar{\bar{X}} = \bar{\bar{Y}}$. We observed, in the proof of H.4, that we then have

$$\begin{aligned}(S_4(\omega_1))([abba]) &= (\ell-k)\,\Sigma(x_i-i)\,(k-i) \\ &\quad - \tfrac{1}{2}\,[(\Sigma(x_i-i))^2 - \Sigma(x_i-i)^2] \\ &= (\ell-k)\,\Sigma(y_i-i)\,(k-i) \\ &\quad - \tfrac{1}{2}\,[(\Sigma(y_i-i))^2 - \Sigma(y_i-i)^2] \\ &= (S_4(\omega_2))([abba]).\ \square\end{aligned}$$

Corollary H.6. If ω_1 and ω_2 are binary strings realizing the same 3-spectrum, then $(S_4(\omega_1))([baab]) = (S_4(\omega_2))([baab])$.

Proof. Let ω_1 and ω_2 be binary strings realizing the same 3-spectrum. By C.11, $S_3(-\omega_1) = S_3(-\omega_2)$; so, by H.5, $(S_4(-\omega_1))([abba]) = (S_4(-\omega_2))([abba])$. But, as baab = - abba, we have

$$\begin{aligned}&(S_4(\omega_1))([baab]) = (S_4(-\omega_1))([abba]) \\ \& \ &(S_4(\omega_2))([baab]) = (S_4(-\omega_2))([abba]).\ \square\end{aligned}$$

Corollary H.7. Let ω_1 and ω_2 be binary strings b-represented by X and Y, respectively. Then $S_4(\omega_1) = S_4(\omega_2)$ iff ω_2 is a rearrangement of ω_1 and

$$\begin{aligned}&\Sigma x_i = \Sigma y_i \ \&\ \Sigma i x_i = \Sigma i y_i \ \&\ \Sigma x_i^2 = \Sigma y_i^2 \\ &\&\ \Sigma i^2 x_i = \Sigma i^2 y_i \ \&\ \Sigma i x_i^2 = \Sigma i y_i^2 \ \&\ \Sigma x_i^3 = \Sigma y_i^3.\end{aligned}$$

Remarks. (i) Given the length of a binary string b-represented by X and given the number of b's contained in the string, then, just six additional numbers (Σx_i, $\Sigma i x_i$, Σx_i^2, $\Sigma i^2 x_i$, $\Sigma i x_i^2$, and Σx_i^3) are needed to entirely characterize the 4-spectrum realized by the string.

(ii) Together with the earlier results B.6, F.1, and G.2, this suggests the following general conjecture:

Let ω_1 and ω_2 be binary strings b-represented by X and Y, respectively, and let $n > 0$. Then $S_n(\omega_1) = S_n(\omega_2)$

iff ω_2 is a rearrangement of ω_1 and $\Sigma i^{\alpha} x_i^{\beta} = \Sigma i^{\alpha} y_i^{\beta}$ whenever $\alpha, \beta \in N$ & $\alpha + \beta < n$.

The messiness and non-uniformity of our proofs so far (of B.6, F.1, G.2, H.7), however, are a block to our seeing how to prove this conjecture. (The conjecture might even be false, of course.)

Proof. By H.4,

$$S_3(\omega_1) = S_3(\omega_2) \;\&\; S_4(\omega_1) = S_4(\omega_2) \Leftrightarrow$$
$$S_3(\omega_1) = S_3(\omega_2) \;\&\; \Sigma i^2 x_i = \Sigma i^2 y_i \;\&\; \Sigma i x_i^2 = \Sigma i y_i^2 \;\&\; \Sigma x_i^3 = \Sigma y_i^3.$$

By Fact 3 and Corollary G.2, we get the desired result from this. □

Let us now try to use H.7 to find strings which are not 4-local.

Lemma H.8. There do not exist distinct real vectors X, Y $\in R^3$ such that $\Sigma x_i = \Sigma y_i$ & $\Sigma i x_i = \Sigma i y_i$ & $\Sigma i^2 x_i = \Sigma i^2 y_i$.

Remark. Notice the similarity of this result to G.3.

Proof. Suppose there are such real vectors X, Y. Let D = X - Y. Then

$$D \cdot V_0^3 = D \cdot V_1^3 = D \cdot V_2^3 = 0.$$

By elementary methods, we see that V_0^3, V_1^3, V_2^3 are linearly independent, so that we must have D = 0; i.e., X = Y. □

Corollary H.9. Any binary string containing exactly three b's is 4-local.

Proof. Suppose ω contains exactly three b's and $S_4(\omega') = S_4(\omega)$. By Fact 3 and B.6, ω' is a rearrangement of ω. Let X, Y $\in R^3$ b-represent ω and ω', respectively. By H.7, $\Sigma x_i = \Sigma y_i$ & $\Sigma i x_i = \Sigma i y_i$ & $\Sigma i^2 x_i = \Sigma i^2 y_i$. By Lemma H.8, then, X = Y. As ω' is a rearrangement of ω with the same b-representation, we must have $\omega' = \omega$. □

Corollary H.10. Any binary string containing no more than three b's is 4-local.

Remark. Notice the similarity to F.2, G.5.

Proof. Immediate, by G.5, Fact 5, and H.9. □

Corollary H.11. If there is a binary string that is not 4-local, then it must contain at least four of each character.

Remarks. (i) This gives us $G(7) \leq 4$ once again. (We already knew this from E.12.)

(ii) This and earlier results in Sections F and G suggest the following conjecture:

> Any binary string that is not n-local contains at least n of each character.

Proof. By H.10, such a string must contain at least four b's. By character-indiscernibility, it must contain at least four a's, too. □

As far as we know so far, we can still optimally hope to find two binary strings, each containing four b's, that realize the same 4-spectrum. (As such strings must be longer than eleven, by H.1, they will each contain at least 12 - 4 = 8 a's.) This leads us to prove the following:

Lemma H.12. Let $U, V \in R^4$ be distinct real vectors with $u_1 \geq v_1$. Let $m = u_1 - v_1$, $X = U + (1-v_1)\, V_0^4$, $Y = V + (1-v_1)\, V_0^4$, $D = Y - X$. Then

$$U, V \in Z^4 \cap P^4 \cap M^4$$
$$\& \ \Sigma u_i = \Sigma v_i \ \& \ \Sigma i u_i = \Sigma i v_i \ \& \ \Sigma u_i^{\,2} = \Sigma v_i^{\,2}$$
$$\& \ \Sigma i^2 u_i = \Sigma i^2 v_i \ \& \ \Sigma i u_i^{\,2} = \Sigma i v_i^{\,2} \ \& \ \Sigma u_i^{\,3} = \Sigma v_i^{\,3}.$$

iff v_1 is a positive integer

$$\& \ D = m\,(-1, 3, -3, 1)$$

& m is a positive integer

& x_3 is an odd integer

& $x_3 > 8m + 1$

& $x_2 = \frac{x_3 + 1}{2} - 2m$

& $x_4 = \frac{3x_3 - 1}{2} - 3m$.

Proof. Clearly,

$U, V \in Z^4 \cap P^4 \cap M^4$

& $\Sigma u_i = \Sigma v_i$ & $\Sigma iu_i = \Sigma iv_i$ & $\Sigma {u_i}^2 = \Sigma {v_i}^2$

& $\Sigma i^2 u_i = \Sigma i^2 v_i$ & $\Sigma i{u_i}^2 = \Sigma i{v_i}^2$ & $\Sigma {u_i}^3 = \Sigma {v_i}^3$

iff v_1 is a positive integer

& $X, Y \in Z^4 \cap P^4 \cap M^4$

& $\Sigma x_i = \Sigma y_i$ & $\Sigma ix_i = \Sigma iy_i$ & $\Sigma {x_i}^2 = \Sigma {y_i}^2$

& $\Sigma i^2 x_i = \Sigma i^2 y_i$ & $\Sigma i{x_i}^2 = \Sigma i{y_i}^2$ & $\Sigma {x_i}^3 = \Sigma {y_i}^3$;

so it suffices to show that

$X, Y \in Z^4 \cap P^4 \cap M^4$

& $\Sigma x_i = \Sigma y_i$ & $\Sigma ix_i = \Sigma iy_i$ & $\Sigma {x_i}^2 = \Sigma {y_i}^2$

& $\Sigma i^2 x_i = \Sigma i^2 y_i$ & $\Sigma i{x_i}^2 = \Sigma i{y_i}^2$ & $\Sigma {x_i}^3 = \Sigma {y_i}^3$

iff $D = m\ (-1, 3, -3, 1)$

& m is a positive integer

& x_3 is an odd integer

& $x_3 > 8m + 1$

& $x_2 = \frac{x_3 + 1}{2} - 2m$

& $x_4 = \frac{3x_3 - 1}{2} - 3m$.

Now,

$\Sigma x_i = \Sigma y_i$ & $\Sigma ix_i = \Sigma iy_i$ & $\Sigma i^2 x_i = \Sigma i^2 y_i$

$\Leftrightarrow D \cdot V_0^4 = D \cdot V_1^4 = D \cdot V_2^4 = 0$

$\Leftrightarrow$ D is some multiple of (-1, 3, -3, 1) (by elementary linear algebra).

Since $d_1 = y_1 - x_1 = v_1 - u_1 = -m$, then,

$$\Sigma x_i = \Sigma y_i \;\&\; \Sigma i x_i = \Sigma i y_i \;\&\; \Sigma i^2 x_i = \Sigma i^2 y_i$$

$$\Leftrightarrow D = m\,(-1, 3, -3, 1).$$

Also,

$$\Sigma x_i^2 = \Sigma y_i^2 \Leftrightarrow \Sigma x_i^2 = \Sigma (x_i + d_i)^2$$

$$\Leftrightarrow \Sigma d_i x_i = -\frac{1}{2}\Sigma d_i^2,$$

$$\Sigma i x_i^2 = \Sigma i y_i^2 \Leftrightarrow \Sigma i x_i^2 = \Sigma i (x_i + d_i)^2$$

$$\Leftrightarrow \Sigma i d_i x_i = -\frac{1}{2}\Sigma i d_i^2,$$

$$\Sigma x_i^3 = \Sigma y_i^3 \Leftrightarrow \Sigma x_i^3 = \Sigma (x_i + d_i)^3$$

$$\Leftrightarrow \Sigma d_i x_i^2 + \Sigma d_i^2 x_i = -\frac{1}{3}\Sigma d_i^3.$$

Therefore,

$$\Sigma x_i = \Sigma y_i \;\&\; \Sigma i x_i = \Sigma i y_i \;\&\; \Sigma x_i^2 = \Sigma y_i^2$$
$$\&\; \Sigma i^2 x_i = \Sigma i^2 y_i \;\&\; \Sigma i x_i^2 = \Sigma i y_i^2 \;\&\; \Sigma x_i^3 = \Sigma y_i^3$$

$$\Leftrightarrow \quad D = m\,(-1, 3, -3, 1)$$
$$\&\; -x_1 + 3x_2 - 3x_3 + x_4 = -10m$$
$$\&\; -x_1 + 6x_2 - 9x_3 + 4x_4 = -25m$$
$$\&\; -x_1^2 + 3x_2^2 - 3x_3^2 + x_4^2 + m(x_1 + 9x_2 + 9x_3 + x_4) = 0$$

$$\Leftrightarrow \quad D = m\,(-1, 3, -3, 1)$$
$$\&\; x_1 = 3x_3 - 2x_4 - 5m$$
$$\&\; x_2 = 2x_3 - x_4 - 5m$$
$$\&\; -x_1^2 + 3x_2^2 - 3x_3^2 + x_4^2 + m(x_1 + 9x_2 + 9x_3 + x_4) = 0$$

$$\Leftrightarrow \quad D = m(-1, 3, -3, 1)$$
$$\&\; x_1 = 3x_3 - 2x_4 - 5m$$
$$\&\; x_2 = 2x_3 - x_4 - 5m.$$

As $x_1 = y_1 - d_1 = (v_1 + (1-v_1)) - (-m) = 1 + m$,

$$D = m(-1, 3, -3, 1)$$
$$\& \; x_1 = 3x_3 - 2x_4 - 5m$$
$$\& \; x_2 = 2x_3 - x_4 - 5m$$
$$\Leftrightarrow \quad D = m(-1, 3, -3, 1)$$
$$\& \; x_2 = \frac{x_3 + 1}{2} - 2m$$
$$\& \; x_4 = \frac{3x_3 - 1}{2} - 3m.$$

Putting all this together, we see that it remains only to show that

$$X, Y \in Z^4 \cap P^4 \cap M^4$$
$$\& \; D = m(-1, 3, -3, 1)$$
$$\& \; x_2 = \frac{x_3 + 1}{2} - 2m$$
$$\& \; x_4 = \frac{3x_3 - 1}{2} - 3m$$
$$\Leftrightarrow \quad D = m(-1, 3, -3, 1)$$

& m is a positive integer

& x_3 is an odd integer

$$\& \; x_3 > 8m + 1$$
$$\& \; x_2 = \frac{x_3 + 1}{2} - 2m$$
$$\& \; x_4 = \frac{3x_3 - 1}{2} - 3m.$$

It suffices, then, to assume that

$$D = m(-1, 3, -3, 1)$$
$$\& \; x_2 = \frac{x_3 + 1}{2} - 2m$$
$$\& \; x_4 = \frac{3x_3 - 1}{2} - 3m$$

and to show that

$$X, Y \in Z^4 \cap P^4 \cap M^4$$

$\Leftrightarrow$ m is a positive integer

& x_3 is an odd integer

$\& \; x_3 > 8m + 1$

As U and V are distinct, so are X and Y. As $u_1 \geq v_1$, $x_1 \geq y_1$. If $x_1 = y_1$, then we have $m = x_1 - y_1 = 0$, so that X = Y, a contradiction; therefore, $x_1 > y_1$, so that $m = x_1 - y_1$ is positive. It is enough, then, to make the additional assumption that $m > 0$ and to show that

$$X, Y \in Z^4 \cap P^4 \cap M^4$$

$$\Leftrightarrow \quad m \text{ is an integer}$$

$$\& \; x_3 \text{ is an odd integer}$$

$$\& \; x_3 > 8m + 1.$$

Now,

$$X, Y \in M^4 \Leftrightarrow x_1 < x_2 < x_3 < x_4$$

$$\& \; x_1 - m < x_2 + 3m < x_3 - 3m < x_4 + m.$$

Since m is positive, then,

$$X, Y \in M^4 \Leftrightarrow x_1 < x_2 \; \& \; x_2 + 6m < x_3 \; \& \; x_3 < x_4.$$

As

$$x_1 = m + 1$$

$$\& \; x_2 = \frac{x_3 + 1}{2} - 2m$$

$$\& \; x_4 = \frac{3x_3 - 1}{2} - 3m,$$

$$X, Y \in M^4$$

$$\Leftrightarrow \quad m + 1 < \frac{x_3 + 1}{2} - 2m$$

$$\& \; \frac{x_3 + 1}{2} - 2m + 6m < x_3$$

$$\& \; x_3 < \frac{3x_3 - 1}{2} - 3m$$

$$\Leftrightarrow \quad 6m + 1 < x_3$$

$$\& \; 8m + 1 < x_3$$

$$\& \; 6m + 1 < x_3$$

$$\Leftrightarrow \quad 8m + 1 < x_3.$$

Therefore, it suffices to make the additional assumption that $X, Y \in M^4$ and to show that

$$X, Y \in Z^4 \cap P^4$$

$$\Leftrightarrow \quad m \text{ is an integer}$$

$$\& \; x_3 \text{ is an odd integer.}$$

Since $X, Y \in M^4$ and $m > 0$, $x_1 = m + 1 > 0$ & $y_1 = 1 > 0$, so that we have $X, Y \in P^4$; hence, it remains only to see that

$$X, Y \in Z^4$$

$$\Leftrightarrow \quad m \text{ is an integer}$$

$$\& \; x_3 \text{ is an odd integer.}$$

Clearly, if m is an integer and x_3 is an odd integer, then

$$X = (m + 1, \frac{x_3 + 1}{2} - 2m, x_3, \frac{3x_3 - 1}{2} - 3m) \in Z^4$$

$$\& \; Y = X + m(-1, 3, -3, 1) \in Z^4.$$

Conversely, if x_3 is not an odd integer, then $X \notin Z^4$; and, if $X \in Z^4$ and m is not an integer, then $Y \notin Z^4$. □

<u>Examples</u>. If we take $v_1 = 1$ (so that U = X and V = Y) & $m = 1$ & $x_3 = 11$, then we get

$$U = X = (m + 1, \frac{x_3 + 1}{2} - 2m, x_3, \frac{3x_3 - 1}{2} - 3m) = (2, 4, 11, 13)$$

$$\& \quad V = Y = X + m(-1, 3, -3, 1) = (1, 7, 8, 14).$$

And, sure enough,

$$\Sigma u_i = 30 = \Sigma v_i$$

$$\& \; \Sigma i u_i = 95 = \Sigma i v_i$$

$$\& \; \Sigma u_i^2 = 310 = \Sigma v_i^2$$

$$\& \; \Sigma i^2 u_i = 325 = \Sigma i^2 v_i$$

$$\& \; \Sigma i u_i^2 = 1075 = \Sigma i v_i^2$$

$$\& \; \Sigma u_i^3 = 3600 = \Sigma v_i^3.$$

Similarly, $v_1 = 5$ & $m_1 = 1$ & $x_3 = 11$ gives

$$U = (2, 4, 11, 13) + 4V_0^4 = (6, 8, 15, 17)$$

$$\& \ V = (1, 7, 8, 14) + 4V_0^4 = (5, 11, 12, 18).$$

Corollary H.13. If a binary string ω contains exactly four b's, then there is at most one other binary string that realizes the same 4-spectrum as ω.

Remark. Observe the resemblance between this result and G.12. A natural conjecture suggested is that, for $n > 2$, there can be at most two distinct binary strings realizing the same n-spectrum and each containing exactly n many b's.

Proof. Assume there are three binary strings, each containing exactly four b's, that realize the same 4-spectrum. By cancellation, then, we can take distinct binary strings ω_1, ω_2, ω_3 such that $\omega_3(0) = b$. Let $X, Y, Z \in R^4$ b-represent ω_1, ω_2, ω_3, respectively. Without loss of generality, we can assume that $x_1 \geq y_1 \geq z_1 = 1$.

Now, let $m = x_1 - y_1$ & $m' = y_1 - z_1$. By Lemma H.12, m and m′ are positive numbers such that (among other things)

$$y_2 = \frac{y_3 + 1}{2} - 2m'$$

$$\& \ x_2 - m' = \frac{(x_3 - m') + 1}{2} - 2m$$

$$\& \ Y - X = m(-1, 3, -3, 1).$$

As $Y - X = m(-1, 3, -3, 1)$, we have

$$y_2 = x_2 + 3m$$

$$\& \ y_3 = x_3 - 3m;$$

hence,

$$x_2 + 3m = \frac{(x_3 - 3m) + 1}{2} - 2m'$$

$$\& \ x_2 - m' = \frac{(x_3 - m') + 1}{2} - 2m.$$

This gives two expressions for x_2:

$$x_2 = \frac{(x_3 - 3m) + 1}{2} - 2m' - 3m$$

$$x_2 = \frac{(x_3 - m') + 1}{2} - 2m + m'.$$

But setting

$$\frac{(x_3 - 3m) + 1}{2} - 2m' - 3m = \frac{(x_3 - m') + 1}{2} - 2m + m'$$

gives $m = -m'$, contradicting the fact that m and m' are both positive. □

Viewing Lemma H.12 more positively, we finally get the following (perhaps surprising) result:

Corollary H.14. There are binary strings which are not 4-local.

Proof. By H.7, the example above shows that binary strings of the same length b-represented by (2, 4, 11, 13) and (1, 7, 8, 14) realize the same 4-spectrum; e.g.,

$$S_4(\text{ababaaaaaababa}) = S_4(\text{baaaaabbaaaaab}).\ \Box$$

Corollary H.15. $G(14) > 4$.

Proof. Immediate from the proof of H.14, as founteen is the length of the string ababaaaaaababa. □

Furthermore, it is clear from the example above and H.12 that we have found the shortest binary strings that contain just four b's apiece and fail to be 4-local. (The length of the strings we can get from H.12 is bounded from below by u_4 and v_4, which are minimized when we take $v_1 = 1$. Then we have $u_4 = x_4$ & $v_4 = y_4$. If $m \geq 2$, then $x_4 > x_3 \geq 19 > 14$; so $m = 1$ minimizes max $(\{x_4, y_4\})$. With $m = 1$, minimizing x_3 will minimize both

$$x_4 = \frac{3x_3 - 1}{2} - 3,$$

$y_4 = x_4 + 1.$

As we must have x_3 an odd integer greater than nine, $x_3 = 11$ is the best we can do.) If G(12) or G(13) is greater than four, then, that fact must be witnessed by binary strings containing at least five of each (by character-indiscernibility) character. Using our new search technique (as in the proof of H.1), now, we prove a more complete statement on G(12) and G(13).

Theorem H.16. There is no binary string of length thirteen that is not 4-local and that does not have an initial or final substring of length twelve that is not 4-local. Furthermore, the binary strings of length twelve that are not 4-local are precisely baaabbabaaab, abbaabaaabba, and the strings obtained from them by taking opposites and reverses.

Proof. We apply the technique used in the proof of H.1 to show that $G(11) \leq 4$; our failure to show that $G(13) \leq 4$ will, as a by-product, show us which pairs of strings of length thirteen realize the same 4-spectra.

Assume ω_1 and ω_2 are binary strings of length thirteen such that

$$\omega_1(0) = b \neq \omega_2(0) \ \& \ S_4(\omega_1) = S_4(\omega_2).$$

Let X and Y b-represent ω_1 and ω_2, respectively.

The number of 4-reports on a string of length thirteen is

$$\binom{13}{4} = \frac{13!}{4!\ 9!} = 715$$

Now, the number of these that start at the beginning of the string is

$$\binom{12}{3} = \frac{12!}{3!\ 9!} = 220.$$

Similarly, the respective numbers of these that start at the second, third, fourth, fifth, sixth, seventh, eighth, ninth, and tenth positions are 165, 120, 84, 56, 35, 20, 10, 4, and 1. As 220 + 165 + 120 + 84 + 56

+ 35 + 20 + 10 + 4 + 1 = 715, this accounts for all the 4-reports on a string of length thirteen.

$$\text{Let } m = \sum_{\substack{\overline{\overline{\omega}}=4 \\ \omega(0)=b}} (S_4(\omega_1))\ ([\omega]) = \sum_{\substack{\overline{\overline{\omega}}=4 \\ \omega(0)=b}} (S_4(\omega_2))\ ([\omega]).$$

As $\omega_1(0) = b$, $m - 220$ is the sum of the members of some subset of $\{165, 120, 84, 56, 35, 20, 10, 4, 1\}$; and, as $\omega_2(0) \neq b$, m is the sum of the members of some subset of $\{165, 120, 84, 56, 35, 20, 10, 4, 1\}$, too. An exhaustive search for "primitive cases" (see pages 80 - 81 in the proof of H.1) reveals only the following:

(i) $m = 220 = 165 + 35 + 20$;

(ii) $m = 221 = 220 + 1 = 165 + 56$;

(iii) $m = 221 = 220 + 1 = 120 + 56 + 35 + 10$;

(iv) $m = 224 = 220 + 4 = 120 + 84 + 20$;

(v) $m = 231 = 220 + 10 + 1 = 120 + 56 + 35 + 20$;

(vi) $m = 240 = 220 + 20 = 120 + 84 + 35 + 1$;

(vii) $m = 250 = 220 + 20 + 10 = 165 + 84 + 1$;

(viii) $m = 255 = 220 + 35 = 165 + 56 + 20 + 10 + 4$;

(ix) $m = 259 = 220 + 35 + 4 = 165 + 84 + 10$;

(x) $m = 260 = 220 + 35 + 4 + 1 = 120 + 84 + 56$;

(xi) $m = 265 = 220 + 35 + 10 = 120 + 84 + 56 + 4 + 1$;

(xii) $m = 269 = 220 + 35 + 10 + 4 = 165 + 84 + 20$;

(xiii) $m = 275 = 220 + 35 + 20 = 120 + 84 + 56 + 10 + 4 + 1$;

(xiv) $m = 280 = 220 + 56 + 4 = 165 + 84 + 20 + 10 + 1$;

(xv) $m = 285 = 220 + 35 + 20 + 10 = 165 + 120$;

(xvi) $m = 286 = 220 + 56 + 10 = 165 + 120 + 1$;

(xvii) $m = 296 = 220 + 56 + 20 = 165 + 120 + 10 + 1$;

(xviii) $m = 305 \doteq 220 + 84 + 1 = 165 + 120 + 20$;

(xix) $m = 315 = 220 + 56 + 35 + 4 = 165 + 120 + 20 + 10$;

(xx) $m = 324 = 220 + 84 + 20 = 165 + 120 + 35 + 4$;

(xxi) $m = 340 = 220 + 120 = 165 + 84 + 56 + 35$;

(xxii) $m = 340 = 220 + 120 = 165 + 84 + 56 + 20 + 10 + 4 + 1$.

As in the proof of H.1, we now examine each case according to the primitive case from which it is derived.

(i) Assume the primitive case from which ω_1 and ω_2 are derived is given by (i) above. By H.7, $\Sigma x_i^2 = \Sigma y_i^2$; so we must have

$$1^2 + \cdots = 2^2 + 6^2 + 7^2 + \cdots,$$

where the rest of each sum depends on the end of the corresponding string and must be one of the following:

0,
11^2,
12^2,
13^2,
$11^2 + 12^2$,
$11^2 + 13^2$,
$12^2 + 13^2$,
$11^2 + 12^2 + 13^2$.

Hence, $(2^2 + 6^2 + 7^2) - 1^2 = 88$ must be among the following fourteen numbers:

0,
$11^2 = 121$,
$12^2 = 144$,
$13^2 = 169$,
$11^2 + 12^2 = 265$,
$11^2 + 13^2 = 290$,
$12^2 + 13^2 = 313$.
$11^2 + 12^2 + 13^2 = 434$,

$$12^2 - 11^2 = 23,$$
$$13^2 - 11^2 = 48,$$
$$12^2 + 13^2 - 11^2 = 192,$$
$$13^2 - 12^2 = 25,$$
$$11^2 + 13^2 - 12^2 = 146,$$
$$11^2 + 12^2 - 13^2 = 96.$$

As it is not, this case is ruled out.

In the same way, we can rule out ω_1 and ω_2 being derived from any other primitive case except (xvi); for none of the following numbers is in the above compiled list of fourteen differences:

(ii) $(1^2 + 10^2) - (2^2 + 5^2) = 72,$

(iii) $(3^2 + 5^2 + 6^2 + 8^2) - (1^2 + 10^2) = 33,$

(iv) $(1^2 + 9^2) - (3^2 + 4^2 + 7^2) = 8,$

(v) $(1^2 + 8^2 + 10^2) - (3^2 + 5^2 + 6^2 + 7^2) = 46,$

(vi) $(3^2 + 4^2 + 6^2 + 10^2) - (1^2 + 7^2) = 111,$

(vii) $(2^2 + 4^2 + 10^2) - (1^2 + 7^2 + 8^2) = 6,$

(viii) $(2^2 + 5^2 + 7^2 + 8^2 + 9^2) - (1^2 + 6^2) = 186,$

(ix) $(1^2 + 6^2 + 9^2) - (2^2 + 4^2 + 8^2) = 34,$

(x) $(1^2 + 6^2 + 9^2 + 10^2) - (3^2 + 4^2 + 5^2) = 168,$

(xi) $(3^2 + 4^2 + 5^2 + 9^2 + 10^2) - (1^2 + 6^2 + 8^2) = 130,$

(xii) $(1^2 + 6^2 + 8^2 + 9^2) - (2^2 + 4^2 + 7^2) = 113,$

(xiii) $(3^2 + 4^2 + 5^2 + 8^2 + 9^2 + 10^2) - (1^2 + 6^2 + 7^2) = 209,$

(xiv) $(2^2 + 4^2 + 7^2 + 8^2 + 10^2) - (1^2 + 5^2 + 9^2) = 126,$

(xv) $(1^2 + 6^2 + 7^2 + 8^2) - (2^2 + 3^2) = 137,$

(xvii) $(2^2 + 3^2 + 8^2 + 10^2) - (1^2 + 5^2 + 7^2) = 102,$

(xviii) $(1^2 + 4^2 + 10^2) - (2^2 + 3^2 + 7^2) = 55,$

(xix) $(1^2 + 5^2 + 6^2 + 9^2) - (2^2 + 3^2 + 7^2 + 8^2) = 17,$

(xx) $(2^2 + 3^2 + 6^2 + 9^2) - (1^2 + 4^2 + 7^2) = 64,$

(xxi) $(2^2 + 4^2 + 5^2 + 6^2) - (1^2 + 3^2) = 71$,

(xxii) $(2^2 + 4^2 + 5^2 + 7^2 + 8^2 + 9^2 + 10^2) - (1^2 + 3^2) = 329$.

In the only remaining cases, ω_1 and ω_2 are derived from primitive case (xvi). As $(2^2 + 3^2 + 10^2) - (1^2 + 5^2 + 8^2) = 23$ is in the list of differences solely by virtue of the facts that

$$12^2 - 11^2 = (12^2 + 13^2) - (11^2 + 13^2) = 23,$$

we know the following:

1. $\omega_1(i) = b$ & $\omega_2(i) = a$ for $i = 0, 4, 7, 11$;
2. $\omega_1(i) = a$ & $\omega_2(i) = b$ for $i = 1, 2, 9, 10$;
3. $\omega_1(i) = \omega_2(i)$ for $i = 3, 5, 6, 8, 12$.

Since $\omega_1(12) = \omega_2(12)$, we can already conclude that any two binary strings of length thirteen that realize the same 4-spectrum either begin or end with the same character; hence, by cancellation, any binary string of length thirteen that fails to be 4-local has an initial or final substring of length twelve that fails to be 4-local. Also by cancellation, then, if ω_1' and ω_2' are distinct binary strings of length twelve such that $\omega_1'(0) = b$ & $S_4(\omega_1') = S_4(\omega_2')$, then we know the following:

1′. $\omega_1'(i) = b$ & $\omega_2'(i) = a$ for $i = 0, 4, 7, 11$;

2′. $\omega_1'(i) = a$ & $\omega_2'(i) = b$ for $i = 1, 2, 9, 10$;

3′. $\omega_1'(i) = \omega_2'(i)$ for $i = 3, 5, 6, 8$.

Now, let $k = (S_4(\omega_1'))([b]) = (S_4(\omega_2'))([b])$, and let $U, V \in R^k$ b-represent ω_1' and ω_2', respectively. Assume, for now, that we have $k \leq 6$ & $\frac{1}{k}\Sigma u_i \leq \frac{13}{2}$; i.e., assume that each of our strings contains no more than six (half of twelve) b's and that the average b-position of each of our strings is not "right of center." We know that the b-positions of ω_1' are 1, 5, 8, 12 and possibly some of 4, 6, 7, 9. As $k \leq 6$ & $\frac{1}{k}\Sigma u_i \leq \frac{13}{2}$, then, we must have one of the following seven cases:

(i) $U = (1, 4, 5, 6, 8, 12)$,
$V = (2, 3, 4, 6, 10, 11)$;

(ii) $U = (1, 4, 5, 7, 8, 12)$,
$V = (2, 3, 4, 7, 10, 11)$;

(iii) $U = (1, 4, 5, 8, 9, 12)$,
$V = (2, 3, 4, 9, 10, 11)$;

(iv) $U = (1, 5, 6, 7, 8, 12)$,
$V = (2, 3, 6, 7, 10, 11)$;

(v) $U = (1, 4, 5, 8, 12)$,
$V = (2, 3, 4, 10, 11)$;

(vi) $U = (1, 5, 6, 8, 12)$,
$V = (2, 3, 6, 10, 11)$;

(vii) $U = (1, 5, 8, 12)$,
$V = (2, 3, 10, 11)$.

Now we use H.7 to further investigate these cases. As

$$1 + 5 + 8 + 12 = 26 = 2 + 3 + 10 + 11$$
$$\& \ 1^2 + 5^2 + 8^2 + 12^2 = 234 = 2^2 + 3^2 + 10^2 + 11^2$$
$$\& \ 1^3 + 5^3 + 8^3 + 12^3 = 2366 = 2^3 + 3^3 + 10^3 + 11^3,$$

we have

$$\Sigma u_i = \Sigma v_i \ \& \ \Sigma u_i^2 = \Sigma v_i^2 \ \& \ \Sigma u_i^3 = \Sigma v_i^3$$

in every one of the seven cases. To say anything about $\Sigma i u_i$, $\Sigma i^2 u_i$, $\Sigma i u_i^2$, etc., we must look at the individual cases.

(i) $\Sigma i u_i - \Sigma i v_i = 136 - 136 = 0$,

but $\Sigma i^2 u_i - \Sigma i^2 v_i = 694 - 696 \neq 0$.

(ii) Clearly, $\Sigma i^2 u_i - \Sigma i^2 v_i$ is the same as in case (i).

(iii) $\Sigma i u_i - \Sigma i v_i = 173 - 172 \neq 0$.

(iv) $\Sigma i u_i - \Sigma i v_i = 123 - 124 \neq 0$.

(v) $\Sigma i u_i - \Sigma i v_i = 116 - 115 \neq 0$.

(vi) $\Sigma i u_i - \Sigma i v_i = 103 - 103 = 0$

$\& \ \Sigma i^2 u_i - \Sigma i^2 v_i = 449 - 449 = 0$

$\& \ \Sigma i u_i^{\,2} - \Sigma i v_i^{\,2} = 1027 - 1027 = 0.$

(vii) $\Sigma i u_i - \Sigma i v_i = 83 - 82 \neq 0.$

Thus, by H.7 we must have case (vi); conversely, by the same result, we do have $S_4(\omega_1') = S_4(\omega_2')$ if $U = (1, 5, 6, 8, 12)$ & $V = (2, 3, 6, 10, 11)$. Hence, there is exactly one pair ω_1', ω_2' of binary strings of length twelve with

$$\omega_1'(0) = b \ \& \ k \le 6 \ \& \ \frac{1}{k}\Sigma u_i \le \frac{13}{2} \ \& \ S_4(\omega_1') = S_4(\omega_2');$$

and that pair is

ω_1' = baaabbabaaab,

ω_2' = abbaabaaabba.

Now, suppose ω is any binary string of length twelve that fails to be 4-local. We are done when we show that ω can be obtained from either ω_1' or ω_2' just by taking opposites and reverses. As these operations are their own inverses, it suffices to show that either ω_1' or ω_2' can be obtained from ω just by taking opposites and reverses.

Let m be the average b-position of ω, and let $k = (S_1(\omega))([b])$. Take

$$\omega' = \begin{cases} \omega, & \text{if } k \le 6 \ \& \ m \le \frac{13}{2}; \\ \hat{\omega}, & \text{if } k \le 6 \ \& \ m > \frac{13}{2}; \\ -\omega, & \text{if } k > 6 \ \& \ m \le \frac{13}{2}; \\ -\hat{\omega}, & \text{if } k > 6 \ \& \ m > \frac{13}{2}. \end{cases}$$

Obviously, ω' is a binary string of length twelve containing no more than six b's and without its average b-position right of center. (If $m > \frac{13}{2}$, then the average a-position cannot possibly be right of center, too.) By character-indiscernibility and reversibility, $F(\omega') = F(\omega) > 4$; i.e., ω' is not 4-local.

Now, either $\omega'(0) = b$ or $\omega'(0) = a$. If $\omega'(0) = b$, then we have shown that we must have $\omega' = \omega_1'$. If $\omega'(0) = a$, then take $\omega'' \neq \omega'$ such that $S_4(\omega'') = S_4(\omega')$. By H.1, $G(11) \leq 4$; so, by cancellation, we cannot have $\omega''(0) = a$. Therefore, $\omega'' = \omega_1'$, so that we must have $\omega' = \omega_2'$. ⌐

Corollary H.17. There are exactly eight binary strings of length twelve that fail to be 4-local.

Proof. Let ω_1' and ω_2' be as in the previous proof. By H.16, it suffices to show that the closure of $\{\omega_1', \omega_2'\}$ under taking opposites and reverses has exactly eight members. These operations are their own inverses and commute, so this closure is the set

$$\{\omega_1', \omega_2', \hat{\omega}_1', \hat{\omega}_2', -\omega_1', -\omega_2', -\hat{\omega}_1', -\hat{\omega}_2'\}.$$

As ω_1' = baaabbabaaab & ω_2' = abbaabaaabba, it is easy to see that all eight of these strings are distinct. □

Corollary H.18. There are exactly thirty-two binary strings of length thirteen that fail to be 4-local.

Proof. By H.16 and cancellation, the set of binary strings of length thirteen that fail to be 4-local is just

$$\{b\omega, a\omega, \omega b, \omega a \mid \omega \text{ binary } \& \ \bar{\bar{\omega}} = 12 \ \& \ F(\omega) > 4\}.$$

By H.17, this set apparently has thirty-two members; and it is, in fact, easy to see that all thirty-two of these strings are distinct. □

Corollary H.19. $G(12) = G(13) > 4$.

Proof. Immediate from H.16 by cancellation and definition. □

Corollary H.20. $G(12) \leq 5$.

Proof. Suppose $G(12) \nleq 5$. By E.7, $G(12) > 5$; so, by C.20, there are binary strings ω_1 and ω_2 of length twelve such that $\omega_1(0) = b \neq \omega_2(0)$ & $S_5(\omega_1) = S_5(\omega_2)$. By the kind of reasoning in the proof of H.16, we can assume that ω_1 and ω_2 contain no more than six b's apiece and that their average b-positions are not right of center. By Fact 3, $S_4(\omega_1) = S_4(\omega_2)$; so, by H.16,

$$\omega_1 = \text{baaabbabaaab},$$

$$\omega_2 = \text{abbaabaaabba}.$$

The number of 5-reports on a string of length twelve is

$$\binom{12}{5} = \frac{12!}{5!\ 7!} = 792.$$

The number of these that start at the beginning of the string is

$$\binom{11}{4} = \frac{11!}{4!\ 7!} = 330.$$

Similarly, the respective numbers of these that start at the second, third, fourth, fifth, sixth, seventh, and eighth positions are 210, 126, 70, 35, 15, 5, and 1. As 330 + 210 + 126 + 70 + 35 + 15 + 5 + 1 = 792, this accounts for all the 5-reports on a string of length twelve.

Applying these facts gives

$$\sum_{\substack{\overline{\overline{\omega}}=5 \\ \omega(0)=b}} (S_5(\omega_1))\ ([\omega]) = 330 + 35 + 15 + 1 = 381,$$

$$\sum_{\substack{\overline{\overline{\omega}}=5 \\ \omega(0)=b}} (S_5(\omega_2))\ ([\omega]) = 210 + 126 + 15 = 351,$$

contradicting $S_5(\omega_1) = S_5(\omega_2)$. Therefore, $G(12) \leq 5$. □

Corollary H.21. $G(12) = G(13) = 5$.

Proof. Immediate from Corollaries H.19, H.20. □

Finally, now, let us summarize what we know about G through thirteen.

Corollary H.22.

$G(0) = 0$	$G(7) = 4$
$G(1) = 1$	$G(8) = 4$
$G(2) = 2$	$G(9) = 4$
$G(3) = 2$	$G(10) = 4$
$G(4) = 3$	$G(11) = 4$
$G(5) = 3$	$G(12) = 5$
$G(6) = 3$	$G(13) = 5.$

Proof. By H.2, $G(11) = 4$; so Fact 8 and the remark on page 63 give $G(0)$ through $G(11)$. Corollary H.21 gives $G(12)$ and $G(13)$. □

I. Conjectures and Open Questions

Few of our results so far say anything about n-events or F, G in general, but our results on 0-events through 4-events do suggest the following conjectures:

1. For each n, there is a string that fails to be n-local.
(Cf., B.2, B.8, F.3, G.7, H.14)

2. For each n, there are arbitrarily many binary strings realizing the same n-spectrum and each containing exactly $n + 1$ many b's.
(Cf., G.21 especially.)

3. For each n, there is a binary string containing exactly n many b's that fails to be n-local.
(Cf., consequences of G.11, H.12 especially.)

4. For each $n > 2$, there can be at most two distinct binary strings realizing the same n-spectrum and each containing exactly n many b's.
(Cf., G.12, H.13.)

5. For each $n \neq 0$, any binary string containing no more than $n - 1$ many b's is n-local.
(Cf., B.10, F.2, G.5, H.10.)

6. The range of G is N.
(Cf., H.22.)

7. The range of F is $\{x \in N \mid x \neq 0\}$.
(Cf., H.22 and Fact 18.)

8. The intervals between changes in G are increasing.
(Cf., H.22.)

9. Let ω_1 and ω_2 be binary strings b-represented by X and Y, respectively. Then, for $n > 0$, $S_n(\omega_1) = S_n(\omega_2)$ iff ω_2 is a rearrangement of ω_1 and $\Sigma i^{\alpha} x_i^{\beta} = \Sigma i^{\alpha} y_i^{\beta}$ whenever $\alpha + \beta \leq n - 1$.
(Cf., B.6, F.1, G.2, H.7.)

The means to proving or disproving each of conjectures one through eight might very well be through proving conjecture nine or something similar to it, for a uniform algebraic characterization of n-localness seems like a necessary foundation for further work. In its present form, the final conjecture seems quite plausible, but some initial investigation is beginning to sew some seeds of doubt with respect to its truth; it is feared that the conditions given may be necessary but not sufficient when $n > 4$.

In addition to the listed conjectures and some less significant ones that may be suggested by results so far, there are a number of more open questions whose pursuit might yield some results of interest. Not all of the questions raised in Chapter IV have been answered, for example. We can also ask some new questions raised (sometimes rather tangentially) by the investigation to this point; e.g.,

1. Exactly how often and when does G change?

2. What phenomena occur with larger alphabets that do not occur when we deal with just binary strings?

3. Is it interesting to look at just the first positions of reports instead of looking at the entire reports?

4. If we take the n-spectrum realized by a string to give only the n-reports, are our results any different?

These and most other questions not answered explicitly in the investigation have not yet been considered at all, so the investigation is actually far from completed.

VIII. CONCLUDING REMARKS

The value of the work above is twofold:

1. It provides, in Chapters II, III, and IV, a formal framework for a large part of an area that might be called "observational complexity of formal languages." The framework can make it clear where each investigation in its domain fits into the area and to what questions investigations might address themselves.

2. In Chapters V, VI, and VII results are actually reported within the framework referred to in 1.

Practical applications have not been a conscious consideration in any of this work, but it is conceivable that observational complexity, in some form, might eventually actually find applications in some sort of a physical device. For example, the theory might somehow dictate how much of "the world" each parallel (and indistinguishable) sensor must be able to observe for some computer that is to perform some designated class of tasks. (On pages 15-16 in Chapter II we speak as if such applications are indeed the motivation behind specific definitions of $\widetilde{A}$.)

Theoretical and practical considerations might easily provide motivation for variations of the studies that fall within the framework outlined here. We might, for example, want to invent some "cloudier" version of the meta-character \$ to signify "any character from some-limited-class" rather than just "any character." We might possibly find motivation to limit the information in an "n-spectrum" to just "n-reports" rather than allowing "0-reports" "1-reports," ..., "n-reports." When applications get beyond character strings, we might very well want to concoct multidimensional varieties of the theory.

More radical changes might include choosing a basically different repertoire of observation types and basing assignments of complexity on something other than "localness." Or we could define observational complexity to be more of "how many looks" are required to gather sufficient information to decide membership, rathen than "how wide vision" is required.

We know of no other work strictly in the area covered by this thesis, but there is some resemblance to Minsky and Papert's Perceptrons [1], especially when we think in terms of possible variations. Perceptrons, however, is concerned with a two-dimensional domain ("retina") and only a binary alphabet, and the means of computation allowed are severely limited. In a broad sense, though, it is quite clear that the subject of that work is a specific version of "observational complexity of retinal images."

IX. APPENDICES

A. Logical and Set Theoretic Metalanguage

$\vee$ ... or

& ... and

$\neg$ ... not

$\Rightarrow$ ... only if

$\Leftrightarrow$ ... if and only if

iff ... if and only if

$\forall$ ... universal quantifier

$\exists$ ... existential quantifier

$\emptyset$ ... empty set

$\in$ ... is a member of

$\notin$ ... is not a member of

$\subseteq$ ... is a subset of

$\bar{\bar{A}}$ ... the cardinality of A

$A \cup B$ ... union of A and B

$A \cap B$ ... intersection of A and B

$\bigcup A$ ... union of the members of A

(e.g., $\bigcup \{B_1, B_2, B_3\} = B_1 \cup B_2 \cup B_3$)

$\bigcap A$ ... intersection of the members of A

$\{x \mid \varphi(x)\}$ ... set of all x satisfying φ.

B. Ordered Pairs and Functions

An ordered pair is an object with a first component and a second component. We write (x, y) for the ordered pair whose first component is x and whose second component is y.

A partial function φ from A to B is a set of ordered pairs whose first and second components are members of A and B, respectively, and with the property that distinct members of φ have different first components. If $(x, y) \in \varphi$ then we write $\varphi(x) = y$ (or say y is the value of φ at x) and say that φ is defined at x (or that $\varphi(x)$ is defined). If f is a partial function defined on all of A, then we say that $f: A \to B$ is a (total) function, and we call A the domain (domain(f)) of f and $\{f(x) \mid x \in A\}$ the range (range(f)) of f. A function whose range has cardinality one is constant. A function $f : A \to B$ whose range is all of B is called onto, and a function $f: A \to B$ for which $f(x) = f(y)$ only if $x = y$ is called one-one. A function $f : A \to B$ that is one-one and onto is called a bijection from A to B or a one-one correspondence between A and B. If $f : A \to B$ is a bijection then the inverse of f is that function (in fact, bijection) $f^{-1} : B \to A$ defined by

$$f^{-1}(y) = x \Leftrightarrow f(x) = y.$$

A bijection from A to itself is called a permutation of A. The permutation defined by $F(x) = x$ is called the identity.

If f and g are functions, then the composition of g and f is that partial function $f \circ g$ defined by

$$f \circ g = \{(x,y) \mid y = f(g(x))\}.$$

(Clearly, composition is associative but not commutative.)

C. The Non-negative Integers

We utilize the Von Neumann concept of ordinals; we identify the empty set ϕ with the integer 0 and the set $n \cup \{n\}$ with the ordinal $n + 1$ whenever n is a non-negative integer. Each non-negative integer, then, is just the set of non-negative integers less than that integer; e.g., $7 = \{0, 1, 2, 3, 4, 5, 6\}$. Therefore each non-negative integer is its own cardinality, and a finite sequence of numbers can be defined as a function on the finite integer that is the cardinality of that sequence.

If A is a finite subset of the set of non-negative integers N, then we denote the greatest member of A by max(A).

If ℓ is a non-negative integer, then we write $\ell!$ for $\ell (\ell - 1) (\ell - 2) \cdots 1$. (We agree that $0! = 1$.) If ℓ and n are non-negative integers, then we write $\binom{\ell}{n}$ for $\frac{\ell!}{n!\,(\ell - n)!}$.

D. Partitions and Equivalence Relations

A partition of a set A is a set of non-empty subsets of A (called the parts of the partition) that are pairwise disjoint and whose union is all of A.

A binary relation on a set A is a set of ordered pairs of members of A. A binary relation B (on A) is an equivalence relation (on A) if it satisfies the following criteria:

(i) $(x, x) \in B$ for each $x \in A$ (reflexivity),

(ii) $(x, z) \in B$ whenever $(x, y) \in B$ & $(y, z) \in B$ (transitivity),

(iii) $(x, y) \in B$ whenever $(y, x) \in B$ (symmetry).

If B is an equivalence relation on A and x is a member of A, then the equivalence class of x (with respect to B) is the set

$$[x] = \{y \in A \mid (x, y) \in B\},$$

and x is a representative of [x]. (When an equivalence class has but one member, it is natural and often useful to "confuse" the class with the member and to speak of them interchangeably.) Clearly, $(x, y) \in B \Leftrightarrow [x] = [y]$, so that the set of equivalence classes with respect to any equivalence relation on A is a partition of A. Conversely, any partition of A induces an equivalence relation ("... is in the same part as...") on A.

E. Real Vectors

We write R for the set of real numbers, and we write R^n for the set of functions from the non-negative integer n into R. We call the members of R^n real vectors or ordered n-tuples of real numbers. We call the member of R^n that is identically zero the 0-vector, denoted simply by 0.

If X is a real vector, then we call X(i) the $(i + 1)^{st}$ coordinate of X. We often write x_i for X(i). We write the real vector $X = \{(0, x_1), (1, x_2), \ldots, (n-1, x_n)\}$ as $(x_1, x_2, \ldots, x_n)$. (The confusion of ordered pairs of real numbers with ordered 2-tuples of real numbers is intentional and natural; an ordered pair essentially is a function on $2 = \{0, 1\}$.)

If $X = (x_1, x_2, \ldots, x_n) \in R^n$, then the number $|X| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$ is called the length of X.

Suppose $X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_n)$ are members of R^n and $t \in R$. Then

$$X \mp Y \text{ denotes } (x_1 \mp y_1, x_2 \mp y_2, \ldots x_n \mp y_n),$$

$$X \cdot Y \text{ denotes } x_1y_1 + x_2y_2 + \ldots + x_ny_n \sum_{i=1}^{n} x_iy_i,$$

and

$$tX \text{ denotes } (tx_1, tx_2, \ldots, tx_n).$$

If $X_1, X_2, \ldots, X_m$ are members of R^n and $t_1, t_2, \ldots, t_m$ are members of R, then $t_1X_1 + t_2X_2 + \ldots + t_mX_m$ is a linear combination of $X_1, X_2, \ldots, X_m$. It is a non-trivial linear combination if some t_i is different from zero. A member of R^n is linearly dependent on (independent of) some other members of R^n if it is (is not) a non-trivial linear combination of finitely many of those other members of R^n. A collection of real vectors is linearly independent if each of them is linearly independent of the rest.

F. Normal Usages of Letters

Here we describe the most common usages of letters used often in the thesis. Many also occur subscripted or primed.

A ... the alphabet; a set

B, C ... sets

D ... the difference between two real vectors of interest

F, G ... see page 17

L ... a language

M^k ... see page 41

N ... the non-negative integers

P ... an event; a "plane" in R^k; a polynomial

P^k ... see page 41

Q^k ... see page 41

R ... the real numbers

R^k... see page 125

S, S_n ... spectrum and n-spectrum

V, V, W, X, Y, Z ... real vectors

Z^k ... see page 41

a, b ... specific characters

f, g ... functions

i, j ... indexing integers; integers

k ... the number of b's contained in a binary string; an integer

ℓ ... the length of a string; an integer

m ... an integer

n ... a level of observational complexity; an integer

p, q ... complete events

t ... a real number

x, y, z ... variable characters; components of real vectors; integers

ω ... a character string

$ ... see page 13

X. References

[1] Minsky, Marvin and Papert, Seymour (1969), Perceptrons, The MIT Press, Cambridge, Mass.

[2] Niven, Ivan and Zuckerman, Herbert S. (1960), An Introduction to the Theory of Numbers, John Wiley & Sons, New York, pp. 99-100.