Keys with NULL values in a Relational Database

Stephan Braun

University of the German Armed Forces Munich

1. Introduction

In a relational database, a relation r with attribute set R is known to be a set of tuples such that for every attribute $A \in R$ there is, within this tuple, an "atomic" value u_A for the domain A of attribute A, the property "atomic" being usually attributed to a one number, one character string or one date value (the latter being actually a compound value). Besides these values, there is usually an additional "NULL" value (for, in principle, every attribute $A \in R$) meaning that NULL may be replaced by any "proper" value of domain A; this may be visualized by stating that the value u_A of attribute A for tuple r is not yet known but may be filled in later[1)].

Now, a relation r is always considered to be a <u>set</u>, in the mathematical sense of this word, of tuples, i.e. there must never be identical tuples within one relation r[2)]. This causes a complication in the presence of NULL values: what if proper values are substituted for some NULL values such that duplicates occur? Substitution of new values for old values in a tuple is a common operation in a relational database, be it the replacement of NULL by a proper value, of a proper value by a different proper value, or also the replacement of a proper value by NULL. The occurrence of duplicates as a consequence of a substitution is usually excluded using the fact that every relation r with attribute set R must have a key K, which is a subset of R, such that, for every tuple $t \in r$, the partial tuple t[K] - using the usual notation – uniquely identifies t. This implies that that there must be no duplicates of key values which, for any substitution, is checked by the DBMS. In order to simplify this checking, it is usually required that NULL values are disallowed at least within t[K], for every $t \in R$; this is sometimes called the "entity integrity rule". Outside of K, i.e. for the attributes of R–K, substitutions are in no way restricted; the absence of duplicates within the "key set" r[K], i.e. the set of the partial tuples t[K] for all $t \in r$, guarantees that there will never be duplicates within r.

In the following, keys containing NULL values will be discussed. It will be shown that NULL values within t[K] for some tuples $t \in r$ are possible <u>without</u> running the risk of duplicates within r[K]. Thus, the entity integrity rule mentioned before turns out to be sufficient but not necessary for the absence of such duplicates. Only the case of replacement of NULL by a proper value will be considered. Replacement of a proper value by a different proper value may, of course, result in duplicates, and only the DBMS can check whether this happens since it depends on which tuples at present belong to r. Replacement of a proper value by NULL may well lead to duplicates, either directly by e.g. replacing all different proper values of two tuples by NULL or, less directly, by replacing NULL values generated before such that two different tuples become identical. However, it will be shown that sets r[K] of tuples containing NULL values can be constructed such that by <u>no</u> replacement of NULLs by proper values duplicates will ever occur.

---

1) Actually the notation NULL is used also for the definitely different situation that a value for attribute A is impossible for a specific tuple r. A common example for this may be seen within an EMPLOYEE table where the attribute "commission" may have a value only for salesmen.

2) Some older versions of RDBMS (e.g. INFORMIX) seem to allow identical tuples within one relation. However, different tuples are always distinguishable by their internal ROWID.

2. Maximally incompatible sets

Any set r of tuples with attribute set R (and possibly with NULL values) is a subset of the set $\neq^R := \prod_{A \in R} (A \cup \{o\})$, $\prod$ denoting the cartesian product of sets, where A is, as before, the domain of all the (proper) values of attribute A, and o is written instead of the somewhat clumsy NULL. Analoguously, any r[K] for $K \subseteq R$ is a subset of $\neq^K := \prod_{A \in K} (A \cup \{o\})$. For the following, the notation $R := \prod_{A \in R} A$ and $K := \prod_{A \in K} A$ will also be used; R and K are the sets of "fully specified" tuples, i.e. of tuples without NULL values, within $\neq^R$ and $\neq^K$, resp.

Not <u>every</u> subset of $\neq^R$ (or of $\neq^K$) is suited as a relation (or as a key set) since the replacement of NULLs by proper values will, in general, produce duplicates. This fact, namely that not every subset of - in our notation - $\neq^R$ is suited as a relation, has, of course, been noted earlier. E.g. in [BUJOH91][*)] the example of a "problematic" relation

A	B
a	o
a	b

with two attributes A, B (with proper values a, b) is given which leads to duplicates by replacing o by b in the first tuple. These authors argue that such a relation is not well-formed because the two tuples are "comparable" using the ordering $o<a$, $o<b$[3)], the first tuple then being below the second, and they suggest that within a well-formed relation the tuples should be pairwise incomparable.

However, if - as stated before - a set of tuples with NULL values is considered well-formed if no replacement of NULL by a proper value produces duplicates, the notion of "compatibility" of tuples appears to be more appropriate:

<u>Def.2.1</u> Let $u, v \in \neq^R$ with values u_A, v_A (or also o) for the $A \in R$. Then, u and v are <u>compatible</u> $:\Leftrightarrow$ for every $A \in R$ either $u_A = o$ or $v_A = o$ or $u_A = v_A$. ↵

Thus, two tuples are compatible iff proper values for the same attribute are identical. The property "compatible" is wider than "comparable"; e.g. the two tuples (a,o) and (o,b) are compatible but not comparable.

Evidently, two tuples can be made identical by replacement of NULLs by proper values iff they are compatible; e.g., (a,o) and (o,b) may be transformed into (a,b) by replacing the second or the first NULL, resp.[4)]

3) This ordering corresponds to the construct of a "flat lattice" in the theory of denotational semantics of programming languages

4) Using the partial ordering among tuples as defined by $o<a$ for any proper value a, (a,b) is the supremum of (a,o) and (o,b), and such a supremum of two tuples with NULL values exists iff the tuples are compatible.

*) P. Bunemann, A. Jung, A. Ohori: Using power-domains to generalize relational databases. Theor. Comp. Sci. <u>91</u>, 23-55 (1991)

Using this notion of compatibility we can define

Def.2.2 A set $r \subseteq \neq^R$ is a relation $:\Leftrightarrow$ the tuples of r are pairwise incompatible ↲

Note that $\neq^R$ itself is not a relation since it contains the tuple (o,…,o) compatible with any other tuple. A relation $r \subseteq \neq^R$, in particular, does not contain duplicates.

Incompatible tuples do not have a supremum and thus can never be made identical by replacement of NULL by a proper value. In order for two tuples u, v to be incompatible there must be at least one attribute $A \in R$ such that $u_A \neq v_A$, with proper values u_A, v_A. Property "incompatible" is narrower than "incomparable".

In particular, R and K are relations since any two tuples from R or from K have different proper values for at least one attribute $A \in R$ or $A \in K$

The notion of compatibility according to Def.2.1 may be used to formally define the notion of a key $K \subseteq R$ for a relation:

Def.2.3 Let $r \subseteq \neq^R$ be a relation according to Def.2.2, i.e. with pairwise incompatible tuples, and $K \subseteq R$, $K \neq \emptyset$. Define the restriction of r to K (denoted as $r|_K$) as the collection (not the set!) of all the tuples t[K] for the $t \in r$.
Then K is called a key for for r $:\Leftrightarrow$ the tuples of $r|_K$ are also pairwise incompatible. ↲

$r|_K$ in general contains duplicates and/or compatible tuples. In general, $r|_K \neq r[K]$, the latter being a set. K is a key for r if $r|_K = r[K]$, i.e. a set, and this set, furthermore, contains only pairwise incompatible tuples. By Def.2.2, R is always a key for r.

If a set $q \subseteq \neq^R$ is not a relation then there is no key for q since q contains compatible tuples and, therefore, any restriction of q will contain parts of these tuples, and these are also compatible. For a relation $r \subseteq \neq^R$ some $K \subseteq R$ may be keys and some not, as shown by the following example:

Let r =

A	B	C
a	b	c
a	b	c′
o	b′	o

with attribute set R = {A,B,C} and $b' \neq b$, $c' \neq c$

r is a relation, as is easily verified. The subset K = {A, B} of R is not a key because

$r|_K =$ ($\neq r[K]$)

A	B
a	b
a	b
o	b′

contains duplicates. However, K′ = {B, C} is a key because

$r|_{K'} =$ ($= r[K']$)

B	C
b	c
b	c′
b′	o

consists of pairwise incompatible tuples.

Note that there is a NULL value in the third tuple of this key set; however, no replacement of this NULL value by a proper value will produce a duplicate. This illustrates the statement given before that the entity integrity rule is sufficient but not necessary for the absence of duplicates within r[K].

For the following, any subset of $\neq^K$ ($K \subseteq R$, $K \neq \emptyset$) with pairwise incompatible tuples will be called a key set. This extends the narrower usage of the term "key set" as a set of always fully specified tuples in the Introduction.

The example suggests one further notion:

Def.2.4 A key $K \subseteq R$ for a relation r is called "minimal" :⇔ every restriction $r|_N$ with $N \subset K$, $N \neq \emptyset$, contains compatible (or even identical) tuples. (Accordingly, K is not minimal if an $r|_N$ with $N \subset K$ consists of pairwise incompatible tuples[5]) ↵

The set K' of the example above is minimal: we have

$r\|_B =$	B	with identical tuples, and	$r\|_C =$	C	with compatible tuples.
	b			c	
	b			c'	
	b'			o	

In order to check for minimality it is sufficient to consider only those $N \subset K$ which can be obtained from K by omitting one single attribute A since compatible (or identical) tuples within $r|_N$ for these N remain compatible (or identical) by omitting further values. Thus, K is minimal if there are compatible tuples within all $r|_N$ for these particular $N \subset K$.

From the foregoing it appears that sets of pairwise incompatible tuples from $\neq^K$ may be of interest. Now, any such set preserves this property of pairwise incompatibility if some tuples are omitted from it (in general not, of course, if tuples are added!). This leads to the notion of "maximally incompatible" sets:

Def.2.5 A set $M_K \subseteq \neq^K$ of pairwise incompatible tuples is called maximally incompatible (or maximal for short) :⇔ every further tuple $t \in \neq^K$, $t \notin M_K$, is compatible with at least one tuple from M_K. ↵

Using this notion the construction of a relation $r \subseteq \neq^R$ may be now described as follows:

Every $r \subseteq \neq^R$ is obtained by first constructing all maximally incompatible sets $M_K \subseteq \neq^K$ for the $K \subseteq R$, $K \neq \emptyset$, then all sets $k \subseteq M_K$, $k \neq \emptyset$, as key sets, and finally, for $K \subset R$, completing the tuples of such a key set by any values (proper or NULL) for the attributes from $R - K$.

Thus, we do not start with relations and look for keys but start with possible keys and complete them to relations. Indeed, any example of a relation can contain only a limited number of tuples, and possible keys for this example may well be fortuitous, i.e. they can loose this property of being a key by updates to the relation. In practice, therefore, keys are defined semantically using knowledge of the real world

5) A key which is not minimal is called a "superkey" elsewhere

to be described by the relations. This, clearly, is not a good starting point for mathematical investigations.

For the maximally incompatible subsets of $\neq^K$ we have

Lemma 2.6

A set $M_K \subseteq \neq^K$ is maximally incompatible $\Leftrightarrow$ the tuples of M_K completely and disjointly cover the set K.

In this, $\mathrm{K} := \prod_{A \in K} A$ is the set of "fully specified" tuples, and

(1) a tuple $t \in M_K$ covers all all the tuples $m \in \mathrm{K}$ with $t \leq m$ using the ordering relation "$\leq$" mentioned before.

E.g. for $K = \{A, B\}$ with $A = \{a_1, a_2\}$ and $B = \{b_1, b_2\}$ the tuple (a_1, o) covers the tuples (a_1, b_1) and (a_1, b_2) from $\mathrm{K} = \{(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2)\}$

(2) a covering of K by M_K is complete iff every $m \in \mathrm{K}$ is covered by at least one $t \in M_K$.

(3) a covering of K by M_K is disjoint iff every $m \in \mathrm{K}$ is covered by at most one $t \in M_K$ (or, alternatively, different $t \in M_K$ cover disjoint subsets of K)

Proof: Let us write for short M instead of M_K. We use the fact, indicated in the example of (1) above, that $u, v \in \neq^K$ are compatible iff there is a tuple $s \in \mathrm{K}$ with $u \leq s$ and $v \leq s$[6]. Then, compatible $u, v \in \neq^K$ would both cover the same $s \in \mathrm{K}$ which for $u, v \in M$ means that the covering of K by M is not disjoint. Thus, if the covering of K by M is disjoint the tuples of M must be pairwise incompatible. If, furthermore, the covering of K by M is complete any further tuple $z \in \neq^K$ covers an $s \in \mathrm{K}$ which is already covered by some $t \in M$. Then, z is compatible with t which means that M is already maximally incompatible.

The other way round, if the covering of K by M is not disjoint there are compatible $u, v \in M$, and if the covering of K by M is not complete then there is an $s \in \mathrm{K}$ without a $t \in M$ and thus this s may be added to M (covering only itself) without destroying the property of pairwise incompatibility which means that M is not maximal. ↵

A maximally incompatible set $M_K \subseteq \neq^K$, then, in general consists of some tuples with NULL values (each of which covers several tuples of K) and in addition of some tuples of K which cover just themselves. Let L_K be the subset of M_K consisting of the (pairwise incompatible) tuples which contain NULL values (for the following such tuples will be called "NULL-tuples"). These tuples cover a subset $S_K := \{s \in \mathrm{K} \mid \exists\, t \in L_K \text{ with } t \leq s\} \subseteq \mathrm{K}$. Let $V_K := \mathrm{K} - S_K$. This is that part of K which is not covered by L_K; the tuples of V_K are fully specified (i.e. without NULL values). Then $M_K = L_K + V_K$ (where the sign "+" will henceforth denote

6) a proof can be given but will be omitted here

disjoint union). In particular, we can have $S_K = K$, i.e. the NULL-tuples of L_K already cover all of K; in this case, $V_K = \emptyset$ (or $M_K = L_K$). We also can have $L_K = \emptyset$ (and, consequently, $S_K = \emptyset$), thus $V_K = K = M_K$. In general, however, the tuples of M_K will be a mix of NULL-tuples and of fully specified tuples.

A maximally incompatible set $M_K \subseteq \neq^K$ should have a special property:

<u>Def.2.7</u> Let $M_K \subseteq \neq^K$ be a maximally incompatible set of tuples. Then M_K is called "attribute-minimal" :⇔ every restriction $M_K|_N$ for $N \subset K$ contains compatible (or even identical) tuples. ↵
Note that, as already used before, it suffices to consider only $N \subset K$ with just one attribute removed from K.
When constructing a relation r with attribute set R, starting with an $M_K \subseteq \neq^K$ for some $K \subseteq R$ and K the key of r, then K should be minimal in the sense of Def.2.4. If M_K is not attribute-minimal then some $N \subset K$ would be a suitable key for r, i.e. K would not be minimal.
For a special situation we have attribute-minimality of M_K guaranteed:

<u>Lemma 2.8</u> If $V_K \neq \emptyset$ in $M_K = L_K + V_K$ then M_K is attribute-minimal.
<u>Proof</u>: Let us write for short M for M_K, L for L_K, and V for V_K, i.e. $M = L + V$, $V \neq \emptyset$ (and M covers all of K). L consists of NULL-tuples and V of fully specified tuples. Consider what happens when one of the attributes, Z say, is removed from K; let $N := K - \{Z\}$ and $N := \prod_{A \in K - \{Z\}} A$. Then consider $L|_N$; this is the collection of all the tuples which are obtained from L by removing Z (some of these tuples may be fully specified, namely if their NULL value occurred only with Z). Then, for every such Z, i.e. for every N,
<u>either</u> (1) in $L|_N$ there are compatible (or even identical) tuples. Then, since L is a subset of M, there are compatible tuples also in $M|_N$, for these N.
<u>or</u> (2) the tuples of $L|_N$ are pairwise incompatible. Then <u>either</u> (2.1) $L|_N$ does not cover all of N <u>or</u> (2.2) it does.
<u>In case (2.1)</u> there must be a tuple $t_0 \in N$ not covered by some $t \in L|_N$, so this t_0 must be a (fully specified) tuple of $V|_N$ ($\neq \emptyset$, since, by assumption, $V \neq \emptyset$). Now, we may assume that for every attribute $Z \in R$ there are at least two proper Z-values (or $|Z| \geq 2$). Therefore, $t_0 \in V|_N$ must be part of at least <u>two</u> tuples of $V \subseteq K$ which means that that t_0 occurs at least two times when forming $V|_N$. Then, since V is a subset of M, there are identical tuples in $M|_N$ also for these N.
<u>In case (2.2)</u> every tuple of N is already covered by some $t \in L|_N$. This implies $V|_N = \emptyset$ which is impossible because $V \neq \emptyset$ by assumption, i.e. case (2.2) cannot occur. ↵

Considering Lemma 2.8 it appears to be of interest to study the case of $V_K = \emptyset$, i.e. of an L_K (all the tuples of which are NULL-tuples) which already covers all of K. Probably (and indeed so, as will be shown) not every set of pairwise incompatible NULL-tuples with attribute set $K \subseteq R$ will cover all of K. Of course, it would be rather cumbersome to explicitly list all the tuples of K covered by an L_K in order to find out if some tuples of K are missing. However, there is a much simpler possibility by just <u>counting</u> the number of the K-tuples covered by an L_K:

<u>Lemma 2.9</u> Let $n_A := |A| \geq 2$ the number of proper values of attribute A, let z be a tuple from a set L_K of pairwise incompatible NULL-tuples with attribute set K, and let z[A] be the value of A within tuple z, including the case z[A] = NULL. Then the number of K-tuples covered by L_K is

$$n(L_K) = \sum_{z \in L_K} n_z \quad \text{with } n_z = \prod_{z[A]=NULL} n_A.$$

In general, $n(L_K) \leq n(K) := \prod_{A \in K} n_A$, n(K) being the number of tuples of K, and L_K covers all of K iff $n(L_K) = n(K)$.

<u>Proof</u>: for every attribute A with z[A] = NULL tuple z covers exactly n_A tuples of K, namely those which are obtained when replacing NULL in z by every (proper) value of A. For several A with z[A] = NULL this holds independently for each of these A, thus z covers exactly $n_z = \prod_{z[A]=NULL} n_A$ tuples of K. Now, since the $z \in L_K$ are pairwise incompatible the subsets of K covered by different $z \in L_K$ are disjoint, thus $n(L_K)$ is the sum of the n_z for the $z \in L_K$. By the same disjointness no $s \in K$ is covered more than once, therefore $n(L_K) \leq n(K)$ with, evidently, $n(K) := \prod_{A \in K} n_A$. So K is completely covered iff $n(L_K) = n(K)$. ↵

Some examples:

1. $K = \{A, B, C\}$ with $n_A = n_B = n_C = 2$ ("binary" attributes) and

L_K =

A	B	C
0	0	1
0	0	2

with a (hopefully) suggestive notation of 0 for NULL and 1,2 for the two proper values of attribute C.

Then $n(L_K) = n_A \cdot n_B + n_A \cdot n_B = 8 = n_A \cdot n_B \cdot n_C = n(K)$, so L_K covers all of K.

2. K = {A, B, C}, $n_A = n_B = n_C = 3$ ("ternary" attributes) and

L_K = (using the same notation as before)

A	B	C
0	1	1
2	0	2
0	0	3

Then $n(L_K) = n_A + n_B + n_A \cdot n_B = 15 < n_A \cdot n_B \cdot n_C = 27$, so $L|_K$ does not cover all of K (the missing K-tuples are: 121, 122, 131, 132, 221, 231, 312, 321, 322, 331, 332)

An L_K covering all of K is, in general, not attribute-minimal as Example 1 above shows or, slightly less trivial, the following example:

3. K = {A, B, C}, $n_A = n_B = n_C = 2$, L_K =

A	B	C
0	1	1
0	2	1
0	0	2

with $n(L_K) = n_A + n_B + n_A \cdot n_B = 8 = n(K)$.
This L_K is not attribute-minimal because omitting attribute A still leaves a set of pairwise incompatible tuples (which covers B × C; cf. Example 1 above when omitting attributes A and B).

On the other hand, there are attribute-minimal sets L_K of pairwise incompatible NULL-tuples covering all of K:

4. K = {A, B, C}, $n_A = n_B = n_C = 2$ and L_K =

A	B	C
0	1	1
0	2	1
1	0	2
2	0	2

with $n(L_K) = n_A + n_A + n_B + n_B = 8 = n(K)$,
i.e. full covering of K

and with attribute-minimality as is easily verified by omitting a single attribute and checking the rest which in any case contains duplicates and/or identical tuples.

An L_K covering all of K is necessarily maximal in the sense of Def.2.5 because any other tuple added to such an L_K would cover a tuple of K already covered before and would, therefore, be compatible with some tuple of L_K. Thus these L_K are suited to be used as M_K. On the other hand, as will be shown later, not every maximal set of pairwise incompatible NULL-tuples covers all of K.

3. Maximally incompatible sets of tuples with NULL-values

The examples and remarks of the foregoing section suggest that it may be worth wile to study the structure of attribute-minimal maximal sets L_K of pairwise incompatible tuples with attribute set K, each tuple containing at least one NULL-value, and L_K covering all of K. In the following, several statements about such sets L_K will be presented; as will be seen, some technical apparatus will be necessary.

First of all, let us put all the tuples of L_K which have NULL in the same subset $I \subseteq K$ into one "group" G_I; all these tuples have proper values for the attributes of $\tilde{I} := K - I$. The cases $I = \emptyset$ or $I = K$ do not apply because $I = \emptyset$ means that there is no NULL value, and $I = K$ means that the tuple contains <u>only</u> NULL values and is, therefore, compatible with any other tuple. Within one such group G_I the tuples are pairwise incompatible iff they differ in their proper parts. For a pair G_I, G_J of groups we have pairwise incompatibility of all $t \in G_I$, $s \in G_J$ iff there is at least one "distinguishing" attribute X which is proper in G_I <u>and</u> in G_J and has <u>disjoint</u> values within G_I and G_J, resp.[8] As is easily seen, such an X exists iff $\tilde{I} \cap \tilde{J} = \emptyset$ or, alternatively, iff $I \cup J \subset K$. Furthermore, L_K shall be attribute-minimal, and for this the intersection of all the I of an L_K must not be empty because an attribute A which has value NULL in all tuples of L_K evidently can be omitted without violating the pairwise incompatibility within L_K[9]. This means, in particular, that there must be at least two groups.

Thus, we consider collections of non-empty subsets $I \subset K$ such that for any such collection □, $|□| \geq 2$, the following conditions hold

(1) $I \cup J \subset K$ for $I, J \in$ □

(2) $\bigcap_{I \in \square} I = \emptyset$

where (1) is necessary for pairwise incompatibility and (2) for attribute-minimality. It will turn out that these two necessary conditions are <u>not</u> sufficient for the construction of the desired L_K (i.e. some of these collections will <u>not</u> be suitable).

A collection □ meeting the two conditions (1) and (2) can be visualized by a "pattern" in which every group G_I is represented by a row of length $|K|$, with an entry of e.g. "o" for an $A \in K$ if the tuples of G_I have NULL for this attribute A, and an entry of "x" if the tuples of G_I have a proper value for this A. The pattern has at least <u>two</u> rows.

The conditions (1) and (2) from above then translate into

(1) two different rows of the pattern never have a total of $|K|$ entries "o"
(2) no column of the pattern consists entirely of "o"

---

8) which is possible if for every attribute $A \in R$ we have $|A| \geq 2$
9) cf. Examples 1 and 3 above

Example: for |K| = 3 we have

o	x	x
x	o	x

as a possible pattern with two groups,

and

o	x	x
x	o	x
x	x	o

or

o	x	x
x	o	x
o	o	x

are possible patterns with 3 groups each.

The two necessary conditions (1) and (2) considerably restrict the possible collections □, in particular for small |K|. Thus, for |K| = 1 there is no □ (which is intuitively clear because for a key consisting of only one attribute a value NULL is compatible with any other value). For |K| = 2 there is no □, either, because the only pattern with two groups is

o	x
x	o

and this violates (1). Thus, an M_K for |K|=2 necessarily contains some fully specified tuples, and NULL may occur only in one of the two attributes[10]. For |K| = 3 the three patterns given above are the only ones (up to permutation of rows or of columns); there are no patterns with 4 or more groups for |K| = 3 because, as is easily verified, any further group would, together with one of the groups already present, violate condition (1). Even then, the first of the two patterns given above for three groups will turn out to be <u>not</u> suitable because it can be shown that a pattern with this structure can never cover all of K:

<u>Theorem 3.1</u> Let L_K be a set of pairwise incompatible NULL-tuples such that every single attribute A ∈ K has value NULL in at least one of the tuples (for the pattern: for every column there is at least one entry "o"). Then no such L_K can cover all of K.

<u>Proof</u>:

1. L_K can cover all of K only if for any subset P ⊆ K the restriction $L_K|_P$ covers all of P (not necessarily disjointly because there may well be compatible or even identical tuples in $L_K|_P$) because for some tuple p_0 ∈ P not covered by a tuple from $L_K|_P$ no tuple s ∈ K containing this p_0 as a part can be covered by a tuple from L_K.
2. For L_K to consist of pairwise incompatible tuples we need, in particular, for any pair of groups an attribute distinguishing between the tuples of these groups, by having disjoint proper values for both groups. Let □ be a collection of subsets I ⊆ K (meeting (1) and (2) of above), and let an attribute X have proper values only for the tuples of the groups G_I with I ∈ □′ ⊂ □ and value NULL for the tuples of the groups G_I with I ∈ □ – □′ ≠ ∅[11]. Then, attribute X is suitable for distinguishing between tuples of different groups G_I only for I ∈ □′. Now, if every other attribute also has value NULL in at least one group we need

10) This is the case of key K′ of page 3

11) There may be, of course, several groups for which X has value NULL

at least <u>two</u> further attributes, Y and Z say, for the remaining distinctions between tuples of different groups (for short: for distinguishing between different groups) because a single further attribute, Y say, besides X would have to distinguish not only between the groups G_I with $I \in \square - \square'$ (which means that it must have proper values for these G_I) but also for any pair G_I, G_J of groups with $I \in \square'$ and $J \in \square - \square'$ which means that Y must have proper values also for the groups G_I with $I \in \square'$, and thus Y has proper values for <u>all</u> groups contrary to the assumption that every attribute has NULL in at least one group.
Therefore, we need one further attribute besides X and Y, Z say, to take care of the remaining distinctions.

3. Now, assume we use <u>only</u> the three attributes X, Y and Z for <u>all</u> the distinctions between groups. Let X be NULL for a part $\square_1 \subseteq \square$ (more precisely, X has value NULL for exactly the groups G_I with $I \in \square_1$), $\square_1 \neq \varnothing$. Then at least one of the other two attributes must have proper values for all $I \in \square_1$ because otherwise e.g. Y would have proper values only within $\square_Y \subset \square_1$, and NULL within $\square_1 - \square_Y$, so Z must have proper values at least within $\square_1 - \square_Y$ in order to distinguish between the groups there. But if Z has NULL for an $I \in \square_Y$ neither Y nor Z can distinguish between this I and any $J \in \square_1 - \square_Y$ because Y has NULL for J and Z has NULL for I. So let us assume that e.g. Y has proper values for all $I \in \square_1$. Since, by assumption, Y has NULL <u>somewhere</u> within $\square$ this must be within $\tilde{\square}_1 := \square - \square_1$, i.e. Y has NULL for the $I \in \square_2$ with some non-empty $\square_2 \subseteq \tilde{\square}_1$. Now, $\square_2 = \tilde{\square}_1$ is impossible because then Z must distinguish between all I, J with $I \in \tilde{\square}_1$ and $J \in \square_1$, so Z can never have NULL contrary to the assumption. Thus, necessarily $\square_2 \subset \tilde{\square}_1$ which means there is a third nonempty part $\square_3 := \tilde{\square}_1 - \square_2$ of $\square$, with proper values of Y. Thus, we have a partition of $\square$ as $\square = \square_1 + \square_2 + \square_3$. Now, distinctions between an $I \in \square_1$ and a $J \in \square_2$ can be
done neither by X nor by Y because X has NULL within $\square_1$ and Y has NULL within $\square_2$. Thus, in order to have such groups distinguished as well, Z must have proper values in all of $\square_1$ and in all of $\square_2$, so Z must have its NULLs somewhere within $\square_3$. We may assume that Z has NULL in <u>all</u> of $\square_3$ (as is shown later, the possibility for a complete covering will even be less if Z has NULL only in a proper part of $\square_3$). Thus, we arrive at the following situation: X has NULL within $\square_1$ and proper values within $\square_2 + \square_3$,
Y has NULL within $\square_2$ and proper values within $\square_1 + \square_3$, Z has NULL within $\square_3$ and proper values within $\square_1 + \square_2$.

4. We now show that, under these circumstances, none of the possible sets L_P, with P := {X, Y, Z}, of pairwise incompatible tuples can cover all of P even if as many tuples as possible are used: since, as shown before, e.g. Y has NULL within $\square_2$ and Z has NULL within $\square_3$, a group from $\square_3$ and another from $\square_2$ can be distinguished only by X. Let $X_3 \subseteq X$ be the subset of the X-values occurring in groups of $\square_3$

and $X_2 \subseteq X$ the corresponding subset for $\square_2$. Then, necessarily, $X_3 \cap X_2 = \varnothing$ because the same X-value in a tuple of a group of $\square_3$ and in a tuple of a group of $\square_2$ would result in these two tuples being compatible: identical in X, the first one NULL in Z and the second one NULL in Y. Furthermore, within the groups of $\square_3$, and also within the groups of $\square_2$, there should be as many tuples as possible as long as these tuples remain pairwise incompatible. This means that every X-value must occur, either in X_3 or in X_2, because otherwise there would be a tuple with an X-value neither in X_3 nor in X_2 and thus being incompatible with all other tuples of $\square_3$ or of $\square_2$ via this missing X-value. So, because of the desired maximum number of tuples we have a partition $X = X_3 + X_2$ of X. Correspondingly, we have $Y = Y_3 + Y_1$ with Y_3 the subset for $\square_3$ and Y_1 the subset for $\square_1$, and $Z = Z_2 + Z_1$ with Z_2 the subset for $\square_2$ and Z_1 the subset for $\square_1$. Thus, any two groups from different $\square_j$ are distinguished, either by X (for $\square_2$ and $\square_3$) or by Y (for $\square_1$ and $\square_3$) or by Z (for $\square_1$ and $\square_2$). Within each of the $\square_j$ there are two possibilities each for distinguishing groups, for $\square_1$ by Y or by Z, for $\square_2$ by X or by Z, and for $\square_3$ by X or by Y. However, as will be seen presently, the details of this latter distinguishing between groups of the same $\square_j$ are of no relevance. Let us now abbreviate $x_3 := |X_3|$, $x_2 := |X_2|$, $x := |X|$, and correspondingly for Y and Z. Then, within all the groups of e.g. $\square_3$ there can be at most $x_3 \cdot y_3$ tuples. This is the maximum number of pairwise incompatible tuples of L_P with X-value from X_3 and Y-value from Y_3 and Z-value NULL. Correspondingly, there can be at most $x_2 \cdot z_2$ L_P-tuples within $\square_2$, and at most $y_1 \cdot z_1$ L_P-tuples within $\square_1$. Thus, taking care now also of the NULL values within the groups, the tuples of $\square_3$ cover at most $x_3 \cdot y_3 \cdot z$ tuples of P, those of $\square_2$ at most $x_2 \cdot y \cdot z_2$, and those of $\square_1$ at most $x \cdot y_1 \cdot z_1$, with no tuple of P being covered by more than one tuple of L_P. Thus, the maximum number of tuples of P coverable by any L_P is $n(L_P) = x \cdot y_1 \cdot z_1 + x_2 \cdot y \cdot z_2 + x_3 \cdot y_3 \cdot z$. But this number is definitely less than the number $n(P) = x \cdot y \cdot z$ of the tuples of P as is easily verified by replacing $x = x_3+x_2$, $y = y_3+y_1$, and $z = z_2+z_1$ in $n(L_P)$ and in $n(P)$, giving only 6 terms for $n(L_P)$ instead of the 8 terms for $n(P)$. The number of coverable P-tuples would, as indicated above, be even less than the number given as $n(L_P)$ had we chosen a limited number of proper values for Z in $\square_3$ instead of choosing NULL in all of $\square_3$ (note that, as to covering, choosing NULL is equivalent to choosing all the proper values). ↵

Thus, we have an interesting result illustrating the statement given before that the two conditions (1) and (2) for the pattern for groups are necessary but not sufficient. In particular, the pattern

o	x	x
x	o	x
x	x	o

for three attributes and three groups given before, is not suitable.

In looking for patterns, then, we may restrict ourselves to patterns in which at least one attribute has proper values for all groups. For all of the following, we assume that there is exactly one such "everywhere proper" attribute; the case of several such attributes is not investigated. In this situation there are, indeed, suitable L_K covering all of K and, in addition, being attribute-minimal. One such example (already given earlier) is

L_K =

A	B	C
0	1	1
0	2	1
1	0	2
2	0	2

with K = {A, B, C} and A = B = C = {1, 2}

which conforms to the pattern

o	x	x
x	o	x

given already before

with two groups G_A and G_B, with two tuples each. This example illustrates a special type of L_K for the given pattern:

Lemma 3.2

Let X be the attribute with proper values everywhere (i.e. for all groups of the pattern). Then we get a suitable L_K if we distribute a partition of X over the groups and use the full domain Y for any other attribute Y which, within a group, has proper values according to the pattern.

Proof: According to Lemma 2.9, one single tuple z of L_K covers $\prod_{z[A]=NULL} n_A$ tuples of K, with $n_A = |A|$, or, for z within a group G_I, $\prod_{A \in I} n_A$ tuples of K (remember that the attributes A ∈ I are exactly the attributes with z[A] = NULL). If, for one single value of X, the full domain Y is used for every attribute Y with proper values for this group there are $\prod_{Y \in K,\ Y \text{ proper in } G_I} n_Y$ tuples with this single X-value within group G_I, each of which covers (disjointly because all these tuples are incompatible via their Y-values) $\prod_{A \in I} n_A$ tuples of K. Let n_X^I be the number of different X-values within group G_I, then $n_X^I \cdot \prod_{A \in K-X} n_A$ tuples of K are (disjointly) covered by the tuples of G_I (the simple form of the product stemming partially from the NULLs of G_I and partially from using the full domain Y for every "proper" Y except for X). In total, we have $n(L_K) = \sum_{I \in \square}(n_X^I \cdot \prod_{A \in K-X} n_A) = (\sum_{I \in \square} n_X^I) \cdot \prod_{A \in K-X} n_A$ tuples of K covered by L_K, and this becomes the desired maximum value $n(K) = \prod_{A\in K} n_A$ if

(*) $\sum_{I\in\Box} n_X^I = n_X$ holds. But then, using the notation X_I for the subset of X occurring in a group G_I, we have $X_I \cap X_J \neq \emptyset$ for $I \neq J$ because any two groups can be distinguished only by X, and furthermore we must have $\mathrm{N}_{I\in\Box} X_I = X$ because otherwise some value $x_0 \in X - \mathrm{N}_{I\in\Box} X_I \neq \emptyset$ will not be counted in $\sum_{I\in\Box} n_X^I$ leading to $\sum_{I\in\Box} n_X^I < n_X$ which is not (*). So the X_I are a partition of X, i.e. $X = \sum_{I\in\Box} X_I$. ↵

Thus, we get a suitable L_K if we distribute a partition of X over the groups G_I (of course, as seen already in the example above, this does not mean that every X-value occurs only in one tuple). On the other hand, this technique is suited only for patterns having not too many groups because for $n_X = |X|$ as before there can be at most n_X groups each of which, then, has a unique X-value (but consists of more than one tuple). A less trivial partition of X is possible only for $n < n_X$ groups. Let e.g. C = {1,2,3} instead of C = {1, 2} in the example above then some suitable L_K are

A	B	C
0	1	1
0	2	1
0	1	2
0	2	2
1	0	3
2	0	3

or

A	B	C
0	1	1
0	2	1
1	0	2
2	0	2
1	0	3
2	0	3

(and others more);

note that for a given C-value all B-values or all A-values are used.

However, the technique used so far is not the only possibility for constructing an L_K with the desired properties. There may be an attribute X, everywhere proper but without a partition of X over the groups if besides this X another attribute Y is used, not everywhere proper, as the following example shows:

Consider a pattern with four attributes A, B, C, D and four groups

A	B	C	D
o	x	x	x
x	o	x	x
o	o	x	x
x	x	o	x

with binary attributes A, B, D and ternary attribute C, and we use D as the X and C as the Y. Then a suitable L_K, K = {A, B, C, D} is, e.g.,

A	B	C	D
0	1	1	1
0	2	1	1
1	0	2	1
2	0	2	1
0	0	3	1
1	1	0	2
1	2	0	2
2	1	0	2
2	2	0	2

Thus, it appears that there are more things between heaven and earth as it seemed to be so far. This situation, however, deserves separate study, given in the next section.

4. Keys with NULL values and two distinguishing attributes

The example at the end of the last section showed an L_K where a partition of the everywhere proper attribute X (the D of the example) was distributed over only two of the four groups of the given pattern (e.g. the third and the fourth, the value of the first and the second group being the same as that of the third). More technically, let us number the groups as G_1 to G_4 (instead of giving the set $I \subseteq K$ for each of them). Then, for this example, we have $D = D_3 + D_4$ as the partition of $D = \{1, 2\}$, specifically $D_3 = \{1\}$, $D_4 = \{2\}$. Using this numbering further, we may state that for the groups 1 and 2 we have $D_1 = D_3$ and $D_2 = D_3$, so for these groups we have used certain partition elements of the partition of D. As a consequence, attribute D is suited for distinguishing between the groups G_1, G_4 and G_2, G_4 and G_3, G_4 but not within the groups G_1, G_2, G_3. In order to also distinguish between these groups a partition of $C = \{1, 2, 3\}$ is distributed over them. Thus, we can distinguish between any two different groups. Within one group, pairwise incompatibility is obtained by using all possible tuples for the remaining attributes with proper values according to the pattern; in particular, if there are no such remaining attributes, as in group G_3, this group can contain only one tuple. L_K is attribute-minimal and covers all of K.

The situation described above for the everywhere proper attribute D holds, in an extended form, in general:

Lemma 4.1 Let X be the everywhere proper attribute of a given pattern with attribute set K and with m groups, and let, for $i \in \{1,\ldots,m\}$, $X_i \subseteq X$ be the subset of the domain X of X occurring within group G_i. Then, in order to construct an L_K of the type indicated, the following must hold for this attribute X:

(1) There is a subset $R \subseteq \{1,\ldots,m\}$, $|R| \geq 2$, such that the X_i for $i \in R$ form a partition of X, i.e.

$$X = \sum_{i \in R} X_i.$$

(2) let the number set Q be defined as the complement of R with respect to $\{1,\ldots,m\}$, i.e. $Q := \{1,\ldots,m\} - R$. Then for the groups G_j with $j \in Q$ the subset $X_j \subseteq X$ of the X-values of such a group G_j is

$$X_j = \sum_{i \in R_j} X_i$$

where $R_j \subset R$ is a (nonempty) proper subset of the number set R which, for every $j \in Q$, is freely assigned to this j. In other words, the subset X_j for a group G_j with $j \in Q$ is a union of some (but not all!) partition elements of the partition of X from (1).

Proof:

(1) for every group G_i necessarily X_i is a proper subset of X because a group G_i with $X_i = X$ could never be distinguished by X from any other group

and this would, when counting the number of K-tuples covered, behave as if G_i had NULL for X contradicting the assumption of X being the everywhere proper attribute. Furthermore, there must be at least <u>two</u> groups G_ν, G_μ with $X_\nu \cap X_\mu = \emptyset$ because with $X_\nu \cap X_\mu \neq \emptyset$ for all ν, μ X is never suited for any distinction between two groups, so either there is no other everywhere proper attribute besides X which, by Theorem 3.1, means that L_K does not cover all of K, or there is such an attribute which means that this attribute should serve as the X. So, let the groups G_i for the $i \in R \subseteq \{1,\ldots,m\}$ be <u>all</u> the groups with pairwise disjoint X_i ; as already shown, $r := |R| \geq 2$. Then necessarily $\bigcup_{i\in R} X_i = X$, i.e. these X_i for $i \in R$ are a partition of X, because otherwise either, for $r < m$, a group G_μ with $\mu \notin R$ could have, for attribute X, $X_\mu = X_0 := X - \bigcup_{i\in R} X_i \neq \emptyset$ with X_μ disjoint to all X_i with $i \in R$ contradicting the definition of R, or, for $r = m$, we could fill up one group G_i with tuples having as X-value one of the values from X_0 and being, via X, incompatible with all tuples of L_K, i.e. L_K is not maximal. Thus, necessarily $X = \sum_{i\in R} X_i$ for $R \subseteq \{1,\ldots,m\}$ with $|R| \geq 2$.

(2) For $r = m$, i.e. $Q = \emptyset$, we have the case of Lemma 3.2 which is the simpler case already treated in Sec. 3. For $r < m$, the $X_j \subseteq X = \sum_{i\in R} X_i$ for $j \in Q \neq \emptyset$ necessarily overlap with one or more of the partition elements X_i of X for $i \in R$. Let, for every $j \in Q$, $R_j \subseteq R$ be the set of those numbers of partition elements X_i of X with which X_j overlaps; in general, for each of the $j \in Q$ there will be a different $R_j \subseteq R$. Thus, $X_j \subseteq \sum_{i\in R_j} X_i$. However, we even have $X_j = \sum_{i\in R_j} X_i$, i.e. X_j is a union of <u>complete</u> partition elements of X because otherwise G_j could be completed by new tuples which are obtained from the tuples already present in G_j by replacing the X-values of these tuples by X-values from $\sum_{i\in R_j} X_i - X_j \neq \emptyset$, and these tuples would be incompatible with all other tuples of the L_K constructed so far, either, if the distinction is done by X, via the new X-value or, if some other attribute does the distinction, via the incompatibility holding before which, in this case, is not altered by the new X-value. This means, in particular, that the case $R_j = R$ is not possible because then $X_j = \sum_{i\in R} X_i = X$, thus X_j would be no <u>proper</u> subset of X contradicting (1). Therefore, R_j must be a <u>proper</u> subset of R. ↵

Having this result, the example at the end of Sec. 3 can be described as follows: we have four groups G_1 to G_4 (counting, e.g., from top to bottom), with D serving as the X of Lemma 4.1 and with the partition D = {1, 2} = {1} + {2} in, e.g., G_3 and G_4[12], thus $D = D_3 + D_4$ or R = {3, 4}, and, consequently, Q = {1, 2}, with $D_1 = D_3$ and $D_2 = D_3$, or R_1 = {3}, R_2 = {3}. For this example, necessarily every D_j for $j \in Q$ consists of only one partition element of D because |R| = 2, so $|R_j| = 1$ for $R_j \subset R$[13]. This example, thus, is seen to be a special case, namely with the X_j for $j \in Q$ consisting of just one partition element. A more general example is obtained by using a ternary instead of the binary domain of D, e.g. D = {1, 2, 3}, and proceeding as follows: for the same pattern as before let $D = D_1 + D_2 + D_4$ with D_1 = {1}, D_2 = {2}, D_4 = {3}, or R = {1, 2, 4} and Q = {3}, thus the only j to consider is j = 3. Let R_3 = {1, 2} or $D_3 = D_1 + D_2$. Now consider which distinctions can be made by D: this attribute now not only distinguishes within the groups G_1, G_2 and G_4 via the partition of D but also between G_3 and G_4 because D_3 = {1, 2} and D_4 = {3}, thus $D_3 \cap D_4 = \emptyset$. This appears to be an interesting property of the technique used here (which will be discussed in general later): the "primary" attribute (as the everywhere proper attribute will be called in the following) is suited, in general, for more distinctions than just within the groups corresponding to its partition and, of course, no other attribute should be used for these additional distinctions, if possible. Clearly, D cannot make all distinctions. For this example, the missing distinctions, between G_1 and G_3 and between G_2 and G_3, must be made by some other attribute. Suitable for this is attribute C partitioning, e.g., C = {1, 2, 3} as $C = C_2 + C_3$ with C_2 = {1} and C_3 = {2, 3} and using $C_1 = C_2$ (={1}). For the C-values and D-values for the four groups we then obtain

	C	D
G_1	1	1
G_2	1	2
G_3	2,3	1,2
G_4	0	3

which leads to (using the Cartesian product for the C- and D-values and taking the complete domains for A and B) the following L_K :

	A	B	C	D
G_1	0	1	1	1
	0	2	1	1
G_2	1	0	1	2
	2	0	1	2
G_3	0	0	2	1
	0	0	2	2
	0	0	3	1
	0	0	3	2
G_4	1	1	0	3
	1	2	0	3
	2	1	0	3
	2	2	0	3

in which the pairwise incompatibility easily is verified, within one group as well as between different groups. Note that only two of the four attributes of the pattern are used for distinguishing between the different groups.

12) we might as well take G_1 and G_4 , or G_2 and G_4

13) the choice of {3} for both R_j is not necessary; another suitable choice is R_1={3}, R_2={4}

This example, thus, may serve as a (maximal) set of four-attribute key tuples with NULL values; no replacement of NULL by a proper value will produce a duplicate. Furthermore, this L_K covers all of K, as is easily verified, e.g. by counting, and it is attribute-minimal because omitting any single column is seen to produce compatible or even identical tuples. At this point, it should be stressed that attribute-minimality clearly only holds for <u>maximal</u> L_K. Subsets of an L_K are, in general, not attribute-minimal (consider, in particular, just one single group!). This corresponds to the situation in real Relational Databases: a relation with a two-attribute key can have, for some time, pairwise different values for one of the attributes within the key; but this attribute, nevertheless, is not the whole key which will become manifest when inserting certain new tuples.

In the example just given <u>one</u> further attribute C was sufficient for making all the distinctions which the primary attribute D was unable to make. This is always so:

<u>Theorem 4.2</u>
Let, for a pattern with $m \geq 2$ groups, X be the primary attribute with, according to Lemma 4.1, $X = \sum_{i\in R} X_i$ with $R \subset \{1,\ldots,m\}$ and $X_j = \sum_{i\in R_j} X_i$ for $j \in Q = \{1,\ldots,m\} - R$, with $R_j \subset R$. Let $V := \mathrm{N}_{j\in Q} R_j \subseteq R$ be all the numbers from R occurring in some R_j, and let $\tilde{Q}_i := \{j \in Q \mid i \in R_j\} \subseteq Q$ be the set of the $j \in Q$ using a certain $i \in V$ (thus, as R_j is the set of $i \in R$ assigned to a fixed $j \in Q$, $\tilde{Q}_i$ is the inverse assignment, listing those $j \in Q$ to which a fixed $i \in R$ is assigned). Then the following holds:
All distinctions between groups not made by X can be made, having full coverage of K, by only one further attribute Y, not proper everywhere, iff V is a <u>proper</u> subset of R and, for every $i \in V$, we have a partition

$$(*) \quad Y = \sum_{j\in\tilde{Q}_i} Y_j + Y_i \qquad \text{of the domain } Y \text{ of this attribute Y}$$

<u>Proof</u>:
1. Complete coverage of K is possible only if, for $P := \{X, Y\} \subseteq K$, the partial tuples of L_K with respect to P cover all of P. Let us test this covering by counting: Let n_X^ν, n_Y^ν be the number of different values in group G_ν for X or Y, resp., with $n_Y^\nu = n_Y := |Y|$ if G_ν for Y either has NULL or has the complete domain Y; then, G_ν covers $n_X^\nu \cdot n_Y^\nu$ tuples of P, thus in total $n_{XY} := \sum_{\nu=1}^{m} n_X^\nu \cdot n_Y^\nu$ tuples of P are covered. The X-values of some group G_j with $j \in Q$ are $X_j = \sum_{i\in R_j\subseteq V} X_i$, so $n_X^j = \sum_{i\in R_j} n_X^i$ because the X_i are disjoint. Thus, $n_{XY} = \sum_{j\in Q}(\sum_{i\in R_j} n_X^i)\cdot n_Y^j + \sum_{i\in R} n_X^i \cdot n_Y^i$, the latter term stemming from counting the P-tuples covered by the groups G_i with $i \in R$. Now, in general, not all the $i \in R$ are used, within an R_j, for some $j \in Q$; more technically,

let $V := \bigcup_{j\in Q} R_j \subseteq R$ be the set of the $i\in R$ used for the $j\in Q$. With this, we can write $\sum_{i\in R} n_X^i\cdot n_Y^i = \sum_{i\in V} n_X^i\cdot n_Y^i + \sum_{i\in R-V} n_X^i\cdot n_Y^i$ which is not as trivial as it may appear at a first glance because we can reorder the double sum of the first term of n_{XY}:
Let $\tilde{Q}_i := \{j \in Q \mid i \in R_j\} \subseteq Q$, $\tilde{Q}_i \neq \emptyset$, then with $V = \bigcup_{j\in Q} R_j$ we get
$\sum_{j\in Q}(\sum_{i\in R_j} n_X^i)\cdot n_Y^j = \sum_{i\in V} n_X^i\cdot(\sum_{j\in \tilde{Q}_i} n_Y^j)$ which leads to $n_{XY} = \sum_{i\in V} n_X^i\cdot(\sum_{j\in \tilde{Q}_i} n_Y^j) + \sum_{i\in V} n_X^i\cdot n_Y^i + \sum_{i\in R-V} n_X^i\cdot n_Y^i$
$= \sum_{i\in V} n_X^i\cdot(\sum_{j\in \tilde{Q}_i} n_Y^j + n_Y^i) + \sum_{i\in R-V} n_X^i\cdot n_Y^i$ for the number of P-tuples covered. But then, with $\sum_{i\in R} n_X^i = n_X$, the desired maximum number $n_X\cdot n_Y$ of P-tuples covered is obtained iff $n_Y^i = n_Y$ for $i\in R-V$ (i.e. for these groups either Y has NULL or the complete domain Y) and, in addition,

$$(*)\ \sum_{j\in \tilde{Q}_i} n_Y^j + n_Y^i = n_Y$$

holds for all $i\in V$. These latter equations (one for each $i\in V$) now hold iff a <u>partition</u> of Y is distributed over the groups G_j with $j\in \tilde{Q}_i$ together with the group G_i , for every $i\in V$, leading to <u>partition equations</u>

$$(**)\ Y = \sum_{j\in \tilde{Q}_i} Y_j + Y_i$$

for every $i\in V$. (Actually, (**) is not really necessary for (*) because (*) may hold for $\bigcup_{j\in \tilde{Q}_i} Y_j \cup Y_i \subset Y$ if some of these subsets of Y overlap. However, as seen below, some of the necessary distinctions between groups, then, can be made <u>neither</u> by X <u>nor</u> by Y). This means that, in particular, Y has proper values in all groups G_j with $j\in \tilde{Q}_i$, though never the whole domain Y because, according to (**), Y_j must be a proper subset of Y. Now, $\bigcup_{i\in V} \tilde{Q}_i = Q$ because $\bigcup_{i\in V} \tilde{Q}_i \subseteq Q$ by definition and for every $j\in Q$ there is at least <u>one</u> $i\in V$ with $j\in \tilde{Q}_i$.
Thus, attribute Y has a proper subset Y_j of Y for all groups G_j with $j\in Q$ and, by (**), also in all groups G_i with $i\in V$. But this means that for the case $V = R$ attribute Y has proper subsets of Y in all of $Q + R$, i.e. everywhere, thus Y can have NULL somewhere only if V is a <u>proper</u> subset of R which means that not all the partition elements X_i of X may be used when forming the X_j for the $j\in Q$.

Of course, the partition equations (**) should be solvable. This, however, is easily possible (assuming that the domain Y is large enough): even if, as will be the general case, the $\tilde{Q}_i$ for different i overlap and thus the same Y_j appears in several partition equations (for different $i\in V$) the number n_Y can, if sufficiently large, always be broken down into smaller numbers solving (*) for every $i \in V$, namely by choosing the n_Y^j for $j \in \tilde{Q}_i$ sufficiently small so that their sum is less than n_Y, and then filling up to n_Y by n_Y^i. The partitions, then, of Y for the equations (**) are obtained by choosing, for every $i \in V$, appropriate proper subsets of Y with the number of elements obtained for (*). Even in the case of non-overlapping $\tilde{Q}_i$ the subsets of Y for different i may, of course, overlap.

2. Now let us consider which distinctions between which groups are made by which attributes. The assignment of X-values to groups as described makes attribute X to distinguish
   - within the G_i with $i \in R$
   - a G_j with $j \in Q$ from all the G_μ with $\mu \in Q$ and $R_\mu \cap R_j = \emptyset$
   - a G_j with $j \in Q$ from all the G_i with $i \in R$ and $i \notin R_j$, in particular <u>every</u> G_j with $j \in Q$ from all the G_i with $i \in R - V$

   Therefore, X is <u>not</u> able to make distinctions between exactly the following groups:
   - G_j and G_μ with $j,\mu \in Q$ and $R_j \cap R_\mu \neq \emptyset$
   - G_j with $j \in Q$ and G_i with $i \in V$ for $i \in R_j$

   However, all these latter distinctions are made by Y:
   for $j,\mu \in Q$ with $R_j \cap R_\mu \neq \emptyset$ (both being subsets of $V \subset R$) there are common $i \in R_j \cap R_\mu$, thus $j,\mu \in \tilde{Q}_i$ for these i, hence, by (**), $Y_j \cap Y_\mu = \emptyset$ because both partition elements of Y occur in the same partition equation. Thus, Y distinguishes between G_j and G_μ for this case. For $j \in Q$, and $i \in V$ with $i \in R_j$, we have $j \in \tilde{Q}_i$ (assigning i to a j means, of course, assigning j to this i), thus here also Y_j and Y_i occur in the same partition equation, i.e. $Y_j \cap Y_i = \emptyset$ by (**), so Y distinguishes between G_j and G_i also in this case. Thus, Y makes all the distinctions which are not made by X. Note that this would not be the case if some of the subsets of Y occurring within (**) overlap. ↵

The constructions of Theorem 4.2 can be illustrated by the former example (at the bottom of page 17) with four groups: there R = {1, 2, 4}, so Q = {3}, and R_3 = {1, 2} = $V \subset R$, so $\tilde{Q}_1$ = {3} and $\tilde{Q}_2$ = {3}, and R–V = {4}. The two partition equations (with C as the Y) are $C_3 + C_1 = C$ (for i = 1) and $C_3 + C_2 = C$ (for i = 2), both solvable and leading, in particular, to $C_1 = C_2$. All this can easily be seen by looking at the examples. Of course, the theorem does <u>not</u> lead to the $C_1 = C_2$ = {1} of this example and, thus, C_3 = {2, 3}, for C = {1, 2, 3}; a partition $C_1 = C_2$ = {1, 2} and C_3 = {3} would serve as well. Thus, in this example, of the six pairwise distinctions for the four groups attribute X distinguishes (in a shorthand notation) the pairs 12, 14, 24, 34 and Y distinguishes 13 and 23.

In this context, it may be of interest which distinctions are, in general, made by the primary attribute X:

Lemma 4.3 Let X be the primary attribute of a pattern with the properties described in Theorem 4.2, and let, as already used with this theorem, $\tilde{Q}_i = \{j \in Q \mid i \in R_j\}$.
Then the following holds:

X distinguishes between two groups G_ν, G_μ iff ν, μ do not appear in the same set $H_i := \tilde{Q}_i + \{i\}$ (which is the set of the numbers occurring in one partition equation)

Proof: X does not distinguish between G_ν and G_μ iff $X_\nu \cap X_\mu \neq \emptyset$, and the latter is equivalent to $\nu, \mu \in H_i$ for some $i \in V$, which may be seen as follows: for $\nu, \mu \in H_i = \tilde{Q}_i + \{i\}$, $\tilde{Q}_i = \{j \in Q \mid i \in R_j\}$, either ν, μ both within $\tilde{Q}_i \subseteq Q$, so $i \in R_\nu \cap R_\mu$ by definition of $\tilde{Q}_i$, thus $X_i \subseteq X_\nu$ and $X_i \subseteq X_\mu$, i.e. $X_i \subseteq X_\nu \cap X_\mu$ and, therefore, $X_\nu \cap X_\mu \neq \emptyset$ for this case, or (e.g.) $\nu \in \tilde{Q}_i$ and $\mu = i$, i.e. $\mu\ (=i) \in R_\nu$, thus $X_\mu \subseteq X_\nu$ or $X_\nu \cap X_\mu \neq \emptyset$ also for this case.
The other way round, for $X_\nu \cap X_\mu \neq \emptyset$ there is some $i \in R$ with either $i \in R_\nu$ and $i \in R_\mu$, i.e. $\nu, \mu \in \tilde{Q}_i \subseteq H_i$, thus $\nu, \mu \in H_i$; or (e.g.) $\nu \in \tilde{Q}_i$ and $\mu = i$ which means $\nu, \mu \in H_i = \tilde{Q}_i + \{i\}$. ↵

For the example above, $H_1 = \{3, 1\}$ and $H_2 = \{3, 2\}$ which now explains why the pairs 13 and 23 can not be distinguished by X. Note that Lemma 4.3 includes the possibility of the same pair ν, μ occurring in more than one H_i.

This example is, of course, comparatively simple. For less simple (and more interesting) cases the following representation appears to be useful: let a "schema"[14)] be constructed as an array with $|V|$ rows and $|Q| + |V|$ columns such that each H_i for some $i \in V$ is represented by a row containing an entry of (e.g.) "x" at column j if either $j \in \tilde{Q}_i$ or $j = i$, and an empty entry otherwise. To give an example, let, for 8 groups numbered from 1 to 8, R = {5, 6, 7, 8} (and Q = {1, 2, 3, 4}), i.e. $X = X_5 + X_6 + X_7 + X_8$, and let $X_1 = X_2 = X_5 + X_6$, $X_3 = X_5 + X_7$, $X_4 = X_6$ be the X-values for the groups G_j with $j \in Q$, thus $V = \{5, 6, 7\} \subset R$. The schema for this example, then, is

	Q				V		
i ↓ j→	1	2	3	4	5	6	7
$\tilde{Q}_i$ →5	x	x	x		x		
→6	x	x		x		x	
→7			x				x
	↑ R_j	↑	↑	↑			

with its entries given by at first freely deciding on the $R_j \subseteq V$ (with V a proper subset of R!) and filling in the columns of the left part of the schema, then adding the entries in the diagonal of the right part of the schema.

14) not to be confused with the "pattern" used before

The H_i, then, can be immediately read off, using the rows, as

$$H_5 = \{1, 2, 3; 5\} , H_6 = \{1, 2, 4; 6\} , H_7 = \{3; 7\}$$ [15)]

According to Lemma 4.3, the primary attribute X does not distinguish pairs $\nu\mu$ with ν, μ both within H_5 or H_6 or H_7; however, X e.g. distinguishes the pair 34 (and others more) because 3 and 4, though occurring within some H_i, do not occur within the same H_i.

For the remaining distinctions (note that e.g. the pair 12 occurs in more than one H_i) the "number equations" (*) for the "secondary" attribute Y are (abbreviating n_Y as n and n_Y^i as n_i)

$$\begin{array}{ccccc|ccc}
n_1 + & n_2 + & n_3 & & & + n_5 & & = n \\
n_1 + & n_2 & & + n_4 & & & + n_6 & = n \\
& & n_3 & & & & + n_7 & = n
\end{array}$$

with the border between the left and the right part of the schema marked. These equations clearly are solvable by freely choosing values for n_1 to n_4 (although with a sum less than n) and then filling up to n by appropriate values of n_5 to n_7. For each such solution n_1 to n_7 there are, as already mentioned with a former (smaller) example, several solutions of the partition equations for the Y_1 to Y_7 using the n_1 to n_7.

There is one further phenomenon which can be described using this example: in general, some of the distinctions made by X are, in addition, also made by Y (however, there are always distinctions which can be made only by Y). If we, e.g., partition Y as $Y = Y_a + Y_b + Y_c + Y_d$ (we need four partition elements because there are number equations with four numbers on the left hand side) then, e.g.,

$Y_1 = Y_a$, $Y_2 = Y_b$, $Y_3 = Y_c$, $Y_4 = Y_d$ by free choice
and $Y_5 = Y_d$, $Y_6 = Y_c$, $Y_7 = Y_a + Y_b + Y_d$ as a consequence of this choice

solves the (three) partition equations (**). Then, in addition to the distinctions necessarily to be made by Y, Y also distinguishes the pairs 34, 56 and 67 [16)] which, primarily, are distinguished by X.

This finishes the discussion of the case of only two attributes for the distinction between groups, X being the "primary" and Y the "secondary" attribute for these distinctions. In a next step, the case of still one more (a "tertiary") attribute will be discussed; as might be expected, the formal apparatus becomes somewhat more complex for this case.

15) with the semicolon separating the left and the right part of a row
16) this is a consequence of choosing $Y_5 = Y_d$ and $Y_6 = Y_c$ which is a special choice because according to the first two equations only $Y_3 + Y_5 = Y_4 + Y_6$ is required.

5. Keys with NULL values and three distinguishing attributes

Let us now discuss the case of one primary attribute X for some of the distinctions between groups, its ability for distinguishing being restricted by the (free) choice of the form of the X_j for $j \in Q$ or, alternatively, by the number sets H_i as defined and used in the preceding section. Let us, extending the construction of that section, assume that two more attributes, Y and Z say, are used for the distinctions which cannot be made by X, and this, now, in such a way that neither Y nor Z alone makes all these remaining distinctions. For this situation we have

Lemma 5.1 For every $i \in V = \bigcap_{j \in Q} R_j \subset R$ the following holds:

(1) there must be an attribute $A^{(i)}$ for the number set H_i, proper everywhere within H_i and with a proper subset $A_j^{(i)}$ of its domain $A^{(i)}$ for every group G_j with $j \in H_i$.

(2) there is a subset $R^{(i)} \subseteq H_i$ such that $A^{(i)} = \sum_{j \in R^{(i)}} A_j^{(i)}$ is a partition of $A^{(i)}$, $|R^{(i)}| \geq 2$, and for all $j \in Q^{(i)} := H_i - R^{(i)}$ we have $A_j^{(i)} = \sum_{\mu \in R_j^{(i)}} A_\mu^{(i)}$ as a (disjoint) union of partition elements of $A^{(i)}$ with a proper subset $R_j^{(i)}$ of $R^{(i)}$, assigned to $j \in Q^{(i)}$.

Proof: The arguments used in the proofs of Theorem 3.1 and Lemma 4.1 can be repeated here for each of the number sets H_i instead of the number set {1,…,m}, with $A^{(i)}$ for the H_i playing the part of X for {1,…,m}, $R^{(i)} \subseteq H_i$ the part of $R \subseteq \{1,\ldots,m\}$, $Q^{(i)}$ the part of Q and $R_j^{(i)} \subset R^{(i)}$ the part of $R_j \subset R$. In particular, complete covering of P with, now, P = {X, Y, Z} and making all the distinctions within each of the H_i (using the properties already known for the primary attribute X, including its inability for distinguishing within one H_i) can be achieved only if (1) and (2) hold. ↵

Thus, Lemma 5.1. essentially only repeats, for each of the sets H_i, the former construction starting with the set {1,…,m }.

According to Lemma 5.1 there might, in general, be a different attribute $A^{(i)}$ for each $i \in V$. However, as will be argued later, there are only very restricted possibilities for choosing different $A^{(i)}$. Thus, for the following we will take one attribute Y, additional to the primary attribute X, uniformly to be used for all the H_i as the $A^{(i)}$, and called the "secondary" attribute. The treatment in Sec.4, then, handled the case $R^{(i)} = H_i$, or $Q^{(i)} = \emptyset$. Here, therefore, we treat the case $Q^{(i)} \neq \emptyset$ which requires that there must be one further "tertiary" attribute Z serving for those distinctions within each of the H_i which Y does not make. For this tertiary attribute Z, thus, Z_j will be a proper subset of Z for some $j \in H_i$ but not, in general, for all. For these latter $j \in H_i$ it does not make any difference if attribute Z has NULL for the groups G_j or if $Z_j = Z$, i.e. the full domain of Z, since Z is not used for distinguishing with these j, anyway, and with respect to the number of P-tuples covered the effect is the same.

Therefore, $Z_j = Z$ will be assumed for these j and we, thus, henceforth will distinguish between the cases $Z_j = Z$ and $Z_j \subset Z$. The occurrence of at least one NULL for Y, and also for Z, is guaranteed because both attributes have NULL for the G_i with $i \in R - V \neq \emptyset$, i.e. outside the range considered here.

In Sec.4, number equations and partition equations for the secondary attribute were obtained by calculating the number n_{XY} of tuples covered by X and Y and requiring $n_{XY} = n_X \cdot n_Y$. Correspondingly, for three attributes X, Y, Z and requiring $n_{XYZ} = n_X \cdot n_Y \cdot n_Z$ the equations

(1) $n_Y^i = n_Y$ and $n_Z^i = n_Z$ for every $i \in R - V$

(2) $\sum_{j \in H_i} n_Y^j \cdot n_Z^j = n_Y \cdot n_Z$ for every $i \in V$

are obtained. As to the corresponding subsets of Y and Z, resp., for (1) either $Y_i = Y$ or NULL, and also $Z_i = Z$ or NULL, will be suitable. For (2), however, the situation is less trivial than for the case with only one other attribute Y, the partition equations for which were seen to be always solvable. The problem arises iff the H_i for different $i \in V$ overlap because for this case the choice of some Y_j or Z_j for a $j \in H_i$ in order to solve the partition equation for this i influences the solution of some other partition equation(s) for other $i \in V$ since a group G_j for some j can, of course, have only <u>one</u> value set Y_j or Z_j; thus, a certain Y_j or Z_j from one partition equation <u>must</u>, in the case of overlapping H_i, be used also for all other partition equations (which even may, as later examples will show, leave the partition equations unsolvable). For disjoint H_i there is, of course, no such influencing. Overlapping H_i[18] occur iff $X_j = \sum_{i \in R_j} X_i$ with $|R_j| \geq 2$ for at least one $j \in Q$; thus, pairwise disjoint H_i exist iff every X_j for $j \in Q$ consists of just one partition element of X.
By Lemma 5.1 (now substituting Y for $A^{(i)}$) $Y = \sum_{j \in R^{(i)}} Y_j$ for $R^{(i)} \subseteq H_i$, and $Y_j = \sum_{\mu \in R_j^{(i)}} Y_\mu$ for $j \in Q^{(i)} = H_i - R^{(i)}$ with $R_j^{(i)} \subset R^{(i)}$ a number set assigned to $j \in Q^{(i)}$. Using the inverse assignment $\mu \mapsto \tilde{Q}_\mu^{(i)}$ with $\tilde{Q}_\mu^{(i)} := \{j \in Q^{(i)} | \mu \in R_j^{(i)}\} \subseteq Q^{(i)}$ ($\tilde{Q}_\mu^{(i)}$ corresponding to the $\tilde{Q}_i$ of Sec.4) and defining $V^{(i)} := \bigcap_{j \in R^{(i)}} R_j^{(i)}$ (corresponding to the V of Sec.4) and also $H_\mu^{(i)} := \tilde{Q}_\mu^{(i)} + \{\mu\}$ for $\mu \in V^{(i)}$ (corresponding to the $H_i = Q_i + \{i\}$ from Sec.4) the number equation (2) from above takes the form

$$\sum_{\mu \in V^{(i)}} n_Y^\mu \cdot \Big(\sum_{j \in H_\mu^{(i)}} n_Z^j\Big) + \sum_{\mu \in R^{(i)} - V^{(i)}} n_Y^\mu \cdot n_Z^\mu = n_Y \cdot n_Z$$

(corresponding to the result for n_{XY} within the proof of Theorem 4.2) leading,

---

18) more precisely, overlapping $\tilde{Q}_i$ because in different $H_i = \tilde{Q}_i + \{i\}$ the i, of course, differs

because of $Y = \sum_{j\in R^{(i)}} Y_j$, to

(*) $\sum_{j\in H_\mu^{(i)}} n_Z^j = n_Z$ for $\mu \in V^{(i)}$, and $n_Z^\mu = n_Z$ for $\mu \in R^{(i)} - V^{(i)}$

as number equations for Z, and

(**) $\sum_{j\in H_\mu^{(i)}} Z_j = Z$ for $\mu \in V^{(i)}$, and $Z_\mu = Z$ for $\mu \in R^{(i)} - V^{(i)}$

as partition equations for Z, for every $i\in V$.

For a single $i\in V$ these partition equations for Z are always solvable even if the $H_\mu^{(i)}$ (more precisely, the $\tilde{Q}_\mu^{(i)}$) for different $\mu \in V^{(i)}$ overlap (corresponding to the same situation with the partition equations for Y in Sec.4).

However, in the case of overlapping H_i for different $i\in V$ the solutions of these equations for different $i\in V$ interfere, influencing the number equations (2) from above which, as should be stressed, must hold for every $i\in V$ in order to have complete covering. As to distinguishing between the groups, we have (corresponding to Lemma 4.3 and with a proof analogous to that proof):

<u>Lemma 5.2</u> Y distinguishes between two groups G_ν and G_λ with $\nu, \lambda \in H_i$ iff ν and λ do <u>not</u> appear in the same set $H_\mu^{(i)} = \tilde{Q}_\mu^{(i)} + \{\mu\} \subseteq H_i$ for some $\mu \in V^{(i)}$. ↵

Before discussing the general case, some examples appear to be appropriate. First, consider a single $H_i := \{1, 2, 3, 4, 5\}$. Assume $R^{(i)} = \{1, 2, 3\}$, thus $Y = Y_1 + Y_2 + Y_3$ and $Q^{(i)} = \{4, 5\}$, and let $Y_4 = Y_1 + Y_3$ and $Y_5 = Y_3$, thus $V^{(i)} = \{1, 3\}$. The assignment of $j \in Q^{(i)}$ to $R_j^{(i)}$, then, is $4 \mapsto \{1, 3\}$ and $5 \mapsto \{3\}$, thus the inverse assignment is $1 \mapsto \{4\}$ and $3 \mapsto \{4, 5\}$, giving $H_1^{(i)} = \{4; 1\}$ and $H_3^{(i)} = \{4, 5; 3\}$, $\tilde{Q}_1^{(i)} = \{4\}$ and $\tilde{Q}_3^{(i)} = \{4, 5\}$. Using the abbreviations z_μ, z_j , z for n_Z^μ , n_Z^j , n_Z the number equations for Z are (for this i)

(1) $z_2 = z$ for $\mu \in R^{(i)} - V^{(i)} = \{2\}$

and (2)
$$z_1 + z_4 = z \quad \text{for } \mu = 1 \in V^{(i)}$$
$$z_3 + z_4 + z_5 = z \quad \text{for } \mu = 3 \in V^{(i)}$$

leading to $z_1 = z_3 + z_5$. As to the partition equations for Z, this gives $Z_j \subset Z$ for $j \in \{1, 3, 4, 5\}$ and $Z_2 = Z$, with $Z = Z_3 + Z_4 + Z_5$ and $Z_1 = Z_3 + Z_5$. For the other attribute Y, $y = y_1 + y_2 + y_3$, $y_4 = y_1 + y_3$ and $y_5 = y_3$.

The complete number equation for i is, then (with y_j, y for n_Y^j, n_Y,),

$$y_1(z_3 + z_5) + y_2 z + y_3 z + (y_1 + y_3)z_4 + y_3 z_5 = yz$$

which is immediately verified using $y = y_1 + y_2 + y_3$ and $z = z_3 + z_4 + z_5$.

This illustrates the situation for one single $i \in V$. In order to illustrate the solvability for the case of overlapping H_i take the schema (used before, after Lemma 4.3)

		Q				V		
		1	2	3	4 |	5	6	7
	5	x	x	x		x		
V	6	x	x		x		x	
	7			x				x

for 8 groups, with R = {5, 6, 7, 8}, Q = {1, 2, 3, 4} and V = {5, 6, 7}, and with H_5 = {1, 2, 3; 5}, H_6 = {1, 2, 4; 6} , H_7 = {3; 7}.
Now let

- $R^{(5)}$ = {1, 3}, i.e. $Y = Y_1 + Y_3$ and $Q^{(5)}$ = {2, 5}, and let $Y_2 = Y_1$, $Y_5 = Y_1$, thus $R_2^{(5)} = R_5^{(5)}$ = {1} and $V^{(5)}$ = {1}
- $R^{(6)}$ = {1, 6}, i.e. $Y = Y_1 + Y_6$ and $Q^{(6)}$ = {2, 4}, and let $Y_2 = Y_1$, $Y_4 = Y_1$, thus $R_2^{(6)} = R_4^{(6)}$ = {1} and $V^{(6)}$ = {1} with $Y_2 = Y_1$ necessary for i = 6 because the same was defined for i = 5 and, of course, there can be only one definition for Y_2. This illustrates the interference between different i.
  For Y_4 there is no such restriction; defining $Y_4 = Y_6$ would serve as well.
- $R^{(7)}$ = {3, 7}, thus $Q^{(7)} = \emptyset$ and, consequently, $V^{(7)} = \emptyset$ because necessarily $|R^{(i)}| \geq 2$, so for $|H_i| = 2$ (which is the minimum) necessarily $Q^{(i)} = V^{(i)} = \emptyset$.

Using these definitions we obtain $H_1^{(5)}$ = {2, 5; 1} ($\tilde{Q}_1^{(5)}$ = {2, 5}) and $H_1^{(6)}$ = {2, 4; 1} ($\tilde{Q}_1^{(6)}$ = {2, 4}) which overlap, thus influencing the value of Y_j and Z_j for j = 1 and j = 2. The number equations for Z are, then,

for i = 5: $z_3 = z$ from $R^{(5)} - V^{(5)}$ = {3}
$z_1 + z_2 + z_5 = z$ from $H_1^{(5)}$

for i = 6: $z_6 = z$ from $R^{(6)} - V^{(6)}$ = {6}
$z_1 + z_2 + z_4 = z$ from $H_1^{(6)}$

for i = 7: $z_3 = z$, $z_7 = z$ from $R^{(7)} - V^{(7)} = R^{(7)}$ = {3, 7}

leading, in particular, to $z_4 = z_5$

and for Y they are

$y_1 + y_3 = y$ from $R^{(5)}$, $y_1 + y_6 = y$ from $R^{(6)}$ and $y_3 + y_7 = y$ from $R^{(7)}$, leading to $y_6 = y_3$ and $y_7 = y_1$, and, from the definitions for Y_2, Y_4 and Y_5, $y_2 = y_4 = y_5 = y_1$.

The complete number equations $\sum_{j \in H_i} y_j z_j = yz$ for $i \in V = \{5, 6, 7\}$ are, then

i ↓ / j →	1	2	3	4	5	6	7	
5	y_1z_1	$+ y_1z_2$	$+ y_3z$		$+ y_1z_5$			$= yz$
6	y_1z_1	$+ y_1z_2$	$+$	$+ y_1z_5$		$+ y_3z$		$= yz$
7			y_3z				$+ y_1z$	$= yz$

which can all be verified using $y = y_1 + y_3$ and $z = z_1 + z_2 + z_5$. Note that, of course, there may be only <u>one</u> product $y_\nu z_\mu$ in one column of the scheme. It may be of interest to give, for this example, the complete solution, including the values for X, namely

$$X = X_5 + X_6 + X_7 + X_8 \quad \text{from } R = \{5, 6, 7, 8\}$$

and $X_1 = X_2 = X_5 + X_6$, $X_3 = X_5 + X_7$, $X_4 = X_6$

from the columns of the schema. In order to simplify notation, let

$a := X_5$, $b := X_6$, $c := X_7$, $d := X_8$ for X
$a := Y_1$, $b := Y_3$ for Y
$a := Z_1$, $b := Z_2$, $c := Z_5$ for Z

with the evident ambiguousness of a, b, c resolved by the assignment to X, Y, Z. Then the solution for the eight groups can be visualized as

	X	Y	Z
1	a+b	a	a
2	a+b	a	b
3	a+c	b	---
4	b	a	c
5	a	a	c
6	b	b	---
7	c	a	---
8	d	---	---

using "---" instead of NULL (or the complete domain)

The pairwise distinction between the eight groups can be verified from this representation of the solution. In particular, of the 28 distinctions to be made

- X serves within {5, 6, 7, 8} and, of course, for all pairs (j, 8) and, in addition, for the pairs 17, 27, 34, 36, 45 and 47 (which are the pairs not occurring within one H_i).

- Y serves for some (but not all) pairs within the H_i, namely 13, 16 and 37 from the $R^{(i)}$ and, in addition, for the pairs 23, 26, 34, 35, 37, 46, 56, 67 (which are the pairs not occurring within one $H_\mu^{(i)}$)
- Z serves for the remaining pairs 12, 14, 15, 24, 25 within the H_i (which are the pairs occurring within one $H_\mu^{(i)}$)

Note that some of the distinctions made by X are also made by Y but that none of the Y-distinctions are made also by Z. Finally, it should be stressed at this point that the solution given above is, of course, not a three-attribute key but part of a key with more than three attributes, as is already evident by some groups having proper values for all three attributes (cf. the example of Sec.4 with four attributes, only two of which serve for distinguishing between different groups).

For a discussion of the general case of overlapping H_i some more technical apparatus will be needed than used so far. In addition, it will turn out that the solvability of the partition equations for the case of overlapping H_i (which, in turn, are defined by choices for the values of the primary attribute X of the pattern), unfortunately, can not be checked by just looking at the schema and applying some nice criterion. Instead, it is necessary to actually construct all the solutions, if any (which, of course, is a finite problem with low complexity, provided there are not very many groups and/or attributes in the pattern). In particular, as will be demonstrated by suitable examples, for some choices, and even for a whole schema, the partition equations may turn out to be unsolvable; however, this is not predictable in a simple manner.

As a first statement concerning interference we have

<u>Lemma 5.3</u> For $j \in H_i$ $Z_j = Z$ iff $j \in R^{(i)} - V^{(i)}$, and $Z_j \subset Z$ iff $j \in Q^{(i)} + V^{(i)}$
<u>Proof</u>: $H_\mu^{(i)} = \tilde{Q}_\mu^{(i)} + \{\mu\}$ and $\tilde{Q}_\mu^{(i)} \neq \varnothing$, so $|H_\mu^{(i)}| \geq 2$. Therefore, there are at least two Z_j within the left side of (**), so $Z_j \subset Z$ for $j \in H_\mu^{(i)}$ and every $\mu \in V^{(i)}$, thus $Z_j \subset Z$ for $j \in \bigcup_{\mu\in V^{(i)}} H_\mu^{(i)} = \bigcup_{\mu\in V^{(i)}} \tilde{Q}_\mu^{(i)} + V^{(i)} = Q^{(i)} + V^{(i)} \subseteq H_i$. By (**), furthermore, $Z_j = Z$ for $j \in R^{(i)} - V^{(i)}$ which is the rest of H_i . ↵

Thus, the choice of $R^{(i)} \subseteq H_i$ and $V^{(i)} \subseteq R^{(i)}$ for Y decides where the value sets for Z are proper subsets of Z and where not. In particular, for $R^{(i)} = H_i$, thus $Q^{(i)} = \varnothing$, also $V^{(i)} = \varnothing$ because for empty $Q^{(i)}$ no partition elements of Y can be distributed, so $Z_j = Z$ within all of H_i for this case (which means that all the distinctions within H_i are made only by Y). The other extreme, $V^{(i)} = R^{(i)}$, is also possible, in contrast to $V \subset R$ discussed earlier, because $V \subset R$ guarantees that neither Y nor Z have proper values for all groups. Now, for overlapping H_i the result of Lemma 5.3 for one $i \in V$ restricts, in general, the possibilities for other H_i because in the case of $\tilde{Q}_i \cap \tilde{Q}_k \neq \varnothing$ for $i, k \in V$ necessarily <u>either</u> $Z_j = Z$ <u>or</u> $Z_j \subset Z$ for $j \in \tilde{Q}_i \cap \tilde{Q}_k$, so <u>either</u> $j \in R^\lambda - V^\lambda$ <u>or</u> $j \in V^\lambda + Q^\lambda$ for $\lambda \in \{i, k\}$. For the example given before, with $H_5 = \{1, 2, 3, 5\}$ and $H_7 = \{3, 7\}$, necessarily $Z_3 = Z$ because of $|H_7| = 2$, thus $Z_3 = Z$ also for H_5 though

H_5 is much larger than H_3. This illustrates the interference mentioned.

The general case is described by

<u>Theorem 5.4</u> Let $U \subseteq Q$ be the set of the "overlapping" $j \in Q$, i.e. the j which occur in more than one $\tilde{Q}_i$[19]. Let $D \subseteq U$ be the subset of these overlapping $j \in Q$ which, among others, belong also to a two-element H_k , $k \in V$ (they may belong to other, larger, H_i as well), and let $D^{(i)} := D \cap H_i$ be the set of these special overlapping $j \in Q$ belonging to a particular H_i; more precisely, $D^{(i)} = D \cap \tilde{Q}_i$. Then, for every $i \in V$, there is a partition of H_i as $H_i = V^{(i)} + D^{(i)} + F^{(i)} + Q^{(i)}$ (with $R^{(i)} = V^{(i)} + D^{(i)} + F^{(i)}$) with $Y_j \subset Y$ for all the $j \in H_i$, $Z_j = Z$ for $j \in D^{(i)} + F^{(i)}$ ($=R^{(i)} - V^{(i)}$) and $Z_j \subset Z$ for $j \in V^{(i)} + Q^{(i)}$[20].

<u>Proof</u>: For a two-element $H_k = \tilde{Q}_k + \{k\}$ necessarily $R^{(k)} = H_k$ because $|R^{(k)}| \geq 2$, thus $Z_j = Z$ for the single $j \in D^{(k)}$ and also for this $j \in D^{(i)}$ and any $i \in V$. By Lemma 5.3, this means $D^{(i)} \in R^{(i)} - V^{(i)}$, thus not only $V^{(i)} \subseteq R^{(i)}$ but also $D^{(i)} \subseteq R^{(i)}$, with $V^{(i)} \cap D^{(i)} = \emptyset$ because $Z_j \subset Z$ for $j \in V^{(i)}$ by Lemma 5.3. Let $F^{(i)} := R^{(i)} - V^{(i)} - D^{(i)}$ then $F^{(i)} \cap V^{(i)} = F^{(i)} \cap D^{(i)} = \emptyset$, so $R^{(i)} + D^{(i)} + F^{(i)}$ is a partition of $R^{(i)}$. For the $j \in F^{(i)}$, if any, $Z_j = Z$ because, by Lemma 5.3, $Z_j \subset Z$ only for $j \in V^{(i)} + Q^{(i)}$. ↵

The result of this theorem, now, opens a way for constructing solutions for the partition equations for overlapping H_i by partitioning each of the H_i into four disjoint parts $V^{(i)}$, $D^{(i)}$, $F^{(i)}$ and $Q^{(i)}$ although not arbitrarily because necessarily, firstly, $D^{(i)} = D \cap H_i$ is determined by D and H_i which, in turn, are determined by the decisions on the primary attribute X and, furthermore, several other necessary conditions, partly describing dependencies within one H_i, partly dependencies due to Y-values within different H_i , and partly the fact that $Z_j = Z$ and $Z_j \subset Z$ cannot both hold for the same j in different H_i. The details are given in

<u>Lemma 5.5</u>

(1) $V^{(i)} \neq \emptyset \Leftrightarrow Q^{(i)} \neq \emptyset$
(2) $V^{(i)} = R^{(i)} \Rightarrow |Q^{(i)}| \geq 2$
(3) $R^{(i)} \subset R^{(k)}$ is impossible for $i,k \in V$ with $i \neq k$
(4) For overlapping $H_i = R^{(i)}+Q^{(i)}$ and $H_k = R^{(k)}+Q^{(k)}$

(a) for $j \in Q^{(i)} \cap Q^{(k)}$: $Y_j = \sum_{\mu \in R_j^{(i)}} Y_\mu = \sum_{\mu \in R_j^{(k)}} Y_\mu$

(b) for $R^{(i)} \cap R^{(k)} \neq \emptyset$: $\sum_{j \in R^{(i)} - R^{(k)}} Y_j = \sum_{j \in R^{(k)} - R^{(i)}} Y_j$

(c) for $j \in Q^{(i)} \cap R^{(k)}$: $R_j^{(i)} \cap (R^{(k)} - \{j\}) = \emptyset$

(5) $(R^{(i)} - V^{(i)}) \cap (Q^{(k)} + V^{(k)}) = \emptyset$ for $i,k \in V$ with $i \neq k$

19) in terms of the schema, there is more than one entry "x" in column j
20) of course, some of these parts of H_i may be empty

Proof:

(1) $V^{(i)} \neq \emptyset$ means that partition elements of Y are used for the Y_j with $j \in Q^{(i)}$, thus there are such j, i.e. $Q^{(i)} \neq \emptyset$, and, for $Q^{(i)} \neq \emptyset$, partition elements of Y with numbers from $R^{(i)}$ are used for the Y_j with $j \in Q^{(i)}$, therefore $V^{(i)} \neq \emptyset$.

(2) for a one-element $Q^{(i)} = \{j_0\}$, $j_0 \in H_i$, all partition elements Y_j with $j \in V^{(i)} \subseteq R^{(i)}$ must be used for $Y_{j_0} = \sum_{j \in V^{(i)}} Y_j$ which in the case of $V^{(i)} = R^{(i)}$ means $Y_{j_0} = \sum_{j \in R^{(i)}} Y_j$ or $Y_{j_0} = Y$ contradicting $Y_j \subset Y$ for all $j \in H_i$.

(3) by definition of $R^{(k)}$, $\sum_{j \in R^{(k)}} Y_j = Y$, so for $R^{(i)} \subset R^{(k)}$ we get $\sum_{j \in R^{(i)}} Y_j \subset Y$ contradicting the definition of $R^{(i)}$.

(4) for overlapping H_i, H_k there can be only one Y_j for every $j \in H_i \cap H_k$. In particular, for $j \in Q^{(i)} \cap Q^{(k)}$ the equalities given are immediately obtained from the two forms of Y_j for this j, and for $R^{(i)} \cap R^{(k)} \neq \emptyset$ omitting the common partition elements of $Y = \sum_{j \in R^{(i)}} Y_j = \sum_{j \in R^{(k)}} Y_j$ gives (b). For $j \in Q^{(i)}$, $Y_j = \sum_{\mu \in R_j^{(i)}} Y_\mu$, thus $Y_j \cap Y_\mu \neq \emptyset$ for all $\mu \in R_j^{(i)}$, whereas for $j \in R^{(k)}$ $Y_j \cap Y_\mu = \emptyset$ for all $\mu \in R^{(k)}$ but j, thus necessarily (c).

(5) $j \in (R^{(i)} - V^{(i)}) \cap (Q^{(k)} + V^{(k)}) \neq \emptyset$ would require simultaneously $Z_j = Z$ and $Z_j \subset Z$, by Lemma 5.3, which is impossible. ↵

The dependencies of (4a) and (4b) between (sums of) partition elements of Y will be called "equalities". There are, in general, further "implicit" equalities between partition elements caused by using the same partition elements for different $j \in Q^{(i)}$; such equalities are, of course, independent of any overlapping.

Using the definitions of Theorem 5.4 the example given before (with eight groups) is described as follows (using D = {3})

i ↓	$V^{(i)}$	$D^{(i)}$	$F^{(i)}$	$Q^{(i)}$	$R^{(i)}$	$R^{(i)} - V^{(i)}$	$Q^{(i)}+V^{(i)}$	H_i
5	1	3	∅	2,5	1,3	3	1,2,5	1,2,3; 5
6	1	∅	6	2,4	1,6	6	1,2,4	1,2,4; 6
7	∅	3	7	∅	3,7	3,7	∅	3; 7

with rather simple $\tilde{Q}_1^{(5)} = Q^{(5)} = \{2, 5\}$, $\tilde{Q}_1^{(6)} = Q^{(6)} = \{2, 4\}$ (and $\tilde{Q}_7 = \emptyset$) because of $|V^{(i)}| = 1$. Conditions (1), (3) and (5) are met, conditions (2) and (4c) are empty, and the equalities are

- from (4a) : $Q^{(5)} \cap Q^{(6)} = \{2\} \Rightarrow Y_2 = Y_1$ will do the job
- from (4b) : $R^{(5)} \cap R^{(6)} \neq \emptyset \Rightarrow Y_6 = Y_3$ necessarily
  $R^{(5)} \cap R^{(7)} \neq \emptyset \Rightarrow Y_7 = Y_1$ necessarily

In addition, the one-element $V^{(5)} = V^{(6)} = \{1\}$ causes an implicit equality: because of $Y_5 = Y_1$ and $Y_4 = Y_1$ necessarily $Y_5 = Y_4$ (which is not due to the conditions of Lemma 5.5). All these equalities were already mentioned before.

As may be expected, the conditions of Lemma 5.5 restrict, sometimes severely so, the solvability of the partition equations of a given schema. The situation is best illustrated by some examples.

Consider the schema (not used before)

		Q 1	2	3	4	V 5	6	7	8
	5	x	x			x			
V	6		x	x	x		x		
	7	x		x				x	
	8			x					x

with $V = \{5, 6, 7, 8\}$[22] , $Q = \{1, 2, 3, 4\}$
and $H_5 = \{1, 2; 5\}$, $H_6 = \{2, 3, 4; 6\}$, $H_7 = \{1, 3; 7\}$, $H_8 = \{5; 8\}$,

having $U = \{1, 2, 3\}$ and $D = \{3\}$, thus $D^{(i)} = \{3\}$ for $i \in \{6, 7, 8\}$ and $D^{(5)} = \varnothing$.
For this schema, there is no solution with $|V^{(i)}| \geq 2$ for one of the i which can be seen by a detailed discussion (not given here); this stresses the point mentioned before that such a detailed discussion is, indeed, necessary to check for a certain type of solution. There are, however, solutions with $|V^{(i)}| = 1$, to be shown in the following.

Particularly important for the solvability of the partition equations are conditions (4a) and (4b) of Lemma 5.5 because the equalities they cause for the partition elements of Y may well be contradictory.

These conditions are "not critical" if either both are empty which is the case if always $Q^{(i)} \cap Q^{(k)} = R^{(i)} \cap R^{(k)} = \varnothing$ or if the numbers μ, j occurring there are always outside of U so that neither the choice of a Y_μ nor of a Y_j interferes with the solution of some other partition equation. Of course, even for a "critical" case, i.e. if (4a) and/or (4b) are not empty, there may be a solution.

---

22) $R \supseteq V$ is of minor interest here

Example:

Assume, for the schema given before

i ↓	$V^{(i)}$	$D^{(i)}$	$F^{(i)}$	$Q^{(i)}$	$R^{(i)}$	$R^{(i)}- V^{(i)}$	$Q^{(i)}+V^{(i)}$	H_i
5	1	∅	5	2	1,5	5	1,2	1,2; 5
6	2	3	∅	4,6	2,3	3	2,4,6	2,3,4; 6
7	1	3	∅	7	1,3	3	1,7	1,3; 7
8	∅	3	8	∅	3,8	3,8	∅	3; 8

Then, conditions (1), (3) and (5) are met and conditions (2), (4a) are empty. For (4b), $R^{(5)} \cap R^{(7)} \neq \emptyset$ giving $Y_5 = Y_3$, $R^{(6)} \cap R^{(7)} \neq \emptyset$ giving $Y_2 = Y_1$, $R^{(6)} \cap R^{(8)} \neq \emptyset$ giving $Y_8 = Y_2$ $(=Y_1)$, $R^{(7)} \cap R^{(8)} \neq \emptyset$ giving again $Y_8 = Y_1$, so all these equalities are free of contradictions. Condition (4c) is empty except for $Q^{(5)} \cap R^{(6)} = \{2\}$ but $R_2^{(5)} = \{1\}$ (see below) and $R^{(6)} - \{2\} = \{3\}$, thus also (4c) is met. Implicitly, the one-element $V^{(i)}$ lead to $Y_4 = Y_2(=Y_1)$, $Y_6 = Y_2(=Y_1)$, $Y_7 = Y_1$. Thus, in total, $Y = Y_1 + Y_3$, $Y_j = Y_1$ for $j \in \{2, 4, 6, 7, 8\}$ and $Y_5 = Y_3$.

The assignments $j \mapsto R_j^{(i)}$ are

for i = 5: $2 \mapsto \{1\}$
for i = 6: $4 \mapsto \{2\}$, $6 \mapsto \{2\}$
for i = 7: $7 \mapsto \{1\}$

Thus, the inverse assignments $\mu \mapsto \tilde{Q}_\mu^{(i)}$ are

for i = 5: $1 \mapsto \{2\} = \tilde{Q}_1^{(5)}$, so $H_1^{(5)} = \{1, 2\}$ $(= Q^{(5)} + V^{(5)})$
for i = 6: $2 \mapsto \{4, 6\} = \tilde{Q}_2^{(6)}$, so $H_2^{(6)} = \{2, 4, 6\}$ $(= Q^{(6)} + V^{(6)})$
for i = 7: $1 \mapsto \{7\} = \tilde{Q}_1^{(7)}$, so $H_1^{(7)} = \{1, 7\}$ $(= Q^{(7)} + V^{(7)})$

leading to $Z = Z_1 + Z_2 = Z_2 + Z_4 + Z_6 = Z_1 + Z_7$ which is all possible taking $Z_1 := Z_4 + Z_6$, $Z_7 := Z_2$.

The complete solution, then (again in the form of the number equations, with the abbreviations used before), is

i ↓ \ j →	1	2	3	4	5	6	7	8	
5	$y_1(z_4 + z_6)$	$+ y_1z_2$			$+ y_3z$				$= yz$
6		y_1z_2	$+ y_3z$	$+ y_1z_4$		$+ y_1z_6$			$= yz$
7	$y_1(z_4 + z_6)$		$+ y_3z$				$+ y_1z_2$		$= yz$
8			y_3z					$+ y_1z$	$= yz$

which can all be verified using $y = y_1 + y_3$ and $z = z_2 + z_4 + z_6$. Note that, according to Lemma 5.2, all distinctions within one $H_\mu^{(i)}$ are made by Z and all the remaining distinctions within one H_i are made by Y, for all $i \in V$.

On the other hand, there are, as expected, critical cases without a solution. As some examples (not given here) indicate, in such cases when exploiting the equalities for the Y_j eventually either $R^{(i)} \subset R^{(k)}$ for $i, k \in V$ with $i \neq k$ results, contradicting (3) of Lemma 5.5, or $Y_\nu = Y_\lambda$ for $\nu, \lambda \in R^{(i)}$ with $\nu \neq \lambda$ results, contradicting $Y = \sum_{j \in R^{(i)}} Y_j$ as a partition of Y.

Even if all the partition equations for Y are solvable, the partition equations for Z may turn out to be not solvable:

Example: Consider the schema (not used before)

	Q: 1	2	3	4	V: 5	6	7
V 5	x	x			x		
V 6	x		x	x		x	
V 7		x	x				x

with V = {5, 6, 7}, Q = {1, 2, 3, 4}
and $H_5 = \{1, 2; 5\}$, $H_6 = \{1, 3, 4; 6\}$, $H_7 = \{2, 3; 7\}$
having U = {1, 2, 3} and D = ∅ . Assume, for this schema,

i ↓	$V^{(i)}$	$D^{(i)}$	$F^{(i)}$	$Q^{(i)}$	$R^{(i)}$	$R^{(i)}- V^{(i)}$	$Q^{(i)}+V^{(i)}$	H_i
5	1	∅	5	2	1,5	5	1,2	1,2; 5
6	1	∅	4,6	3	1,4,6	4,6	1,3	1,3,4; 6
7	2	∅	7	3	2,7	7	2,3	2,3; 7

meeting conditions (1), (3), (5) of Lemma 5.5, condition (2) being empty. For (4a), $Q^{(6)} \cap Q^{(7)} = \{3\}$ giving $Y_2 = Y_1$ ($=Y_3$), and for (4b) $R^{(5)} \cap R^{(6)} \neq \emptyset$ giving $Y_5 = Y_4 + Y_6$; for (4c), $Q^{(5)} \cap R^{(7)} = \{2\}$ but $R_2^{(5)} = \{1\}$ and $R^{(7)} - \{2\} = \{7\}$, thus (4c) is met. Implicitly, $Y_2 = Y_3 = Y_1$ again. The different partitions of Y are, then, $Y = Y_1 + Y_5 = Y_1 + Y_4 + Y_6 = Y_2 + Y_7$ giving, implicitly, $Y_7 = Y_5$, so, in total, $Y = Y_1 + Y_4 + Y_6$, $Y_2 = Y_3 = Y_1$, $Y_5 = Y_4 + Y_6$, $Y_7 = Y_5$, thus Y makes all the distinctions for ν, λ where ν,λ are not both within one of the $H_\mu^{(i)} = Q^{(i)} + V^{(i)}$, the latter equality holding because $|V^{(i)}| = 1$ for every $i \in V$.

However, the partition equations for Z, then, require

$$Z = Z_1 + Z_2 = Z_1 + Z_3 = Z_2 + Z_3$$

leading to $Z_1 = Z_2 = Z_3$ which clearly contradicts the pairwise distinctness of the elements of a partition of Z.

Examples like those given before support the conjecture that, maybe, there is no simple way of checking the solvability of the partition equations directly for the schema, or for a certain distribution of numbers within the scheme, without actually exploiting the consequences of the different equalitites.

For one of the examples given, it was noted that some distinctions made by the primary attribute X are also made by the secondary attribute Y. More general, the following holds:

<u>Lemma 5.6</u>

Let X be the primary attribute, Y the secondary attribute and Z the tertiary attribute, and let this fact be denoted by $X > Y > Z$, thus introducing the notion of "higher" vs. "lower" attributes. Then

(1) no lower attribute can make <u>all</u> the distinctions of the higher attribute

(2) Z can make <u>none</u> of the distinctions made by Y if always $|V^{(i)}| = 1$.

<u>Proof:</u>

(1) - Y does not distinguish pairs (j, k) with $j \in R - V$ ($\neq \emptyset$) and any $k \in \{1,\ldots,m\}$, $k \neq j$, because of $Y_j = Y$ (or NULL) for these j; X, however, distinguishes all pairs j, k with $j, k \in R$.

- for $V^{(i)} \subset R^{(i)}$ we have the same situation for the tertiary attribute Z: it does not distinguish pairs (j, k) with $j \in R^{(i)} - V^{(i)} \neq \emptyset$ and any $k \in H_i$, $k \neq j$, because of $Z_j = Z$ for these j, by Lemma 5.3; however, Y distinguishes all pairs (j, k) with $j, k \in R^{(i)}$. For $V^{(i)} = R^{(i)}$ (which is a situation not occurring for Y since $V \subset R$), by Lemma 5.3, $Z_j \subset Z$ for all $j \in V^{(i)} + Q^{(i)} = R^{(i)} + Q^{(i)} = H_i$; thus, distinguishing by Z within all of H_i is not a-priori excluded. However, $|R^{(i)}| \geq 2$, thus $|V^{(i)}| \geq 2$, therefore $H_\rho^{(i)} = \tilde{Q}_\rho^{(i)} + \{\rho\}$ and $H_\lambda^{(i)} = \tilde{Q}_\lambda^{(i)} + \{\lambda\}$ for $\rho,\lambda \in V^{(i)}$ with $\rho \neq \lambda$ differ at least by ρ and λ, so, e.g., Z_λ overlaps with some Z_j, $j \in H_\rho^{(i)}$, because of $\sum_{j \in H_\rho^{(i)}} Z_j = \sum_{j \in H_\lambda^{(i)}} Z_j = Z$ by (**), thus Z does not distinguish at least one pair (λ, j) with $j \in H_\rho^{(i)}$ which, by Lemma 5.2, is distinguished by Y because λ, j do not belong to the same set $H_\mu^{(i)}$.

(2) For $|V^{(i)}| = 1$ (for all $i \in V$) always $H_\mu^{(i)} = Q^{(i)} + V^{(i)}$, with $V^{(i)} \subset R^{(i)}$ because of $|R^{(i)}| \geq 2$, thus $Z_j = Z$ for $j \in R^{(i)} - V^{(i)} \neq \emptyset$, by Lemma 5.3, so Z does not distinguish the pairs (j, k) with $j \in R^{(i)} - V^{(i)}$ and any $k \in H_i$. By Lemma 5.2, Y distinguishes a pair (j, k) iff j, k do not appear within the same $H_\mu^{(i)}$, i.e. within $Q^{(i)} + V^{(i)}$, which, here, means that (e.g.) $j \in R^{(i)} - V^{(i)}$ and $k \in H_i$ anywhere. Thus Z does not distinguish exactly the pairs distinguished by Y. ↵

So (2) of this lemma, in particular, explains why in the cited example none of the Y-distinctions were made also by Z. For $|V^{(i)}| \geq 2$, for some $i \in V$, $Z_j \cap Z_k$ might occur for $Z_j \subset Z$ and $Z_k \subset Z$ with $j, k \in Q^{(i)} + V^{(i)}$ but j, k belonging to different $H_\mu^{(i)}$ (for different $\mu \in V^{(i)}$).

This finishes the discussion of three attributes X, Y, Z with X > Y > Z distinguishing between the groups. The evident complexity of the situation for this case indicates that it may not be worth wile to discuss the case of more than three distinguishing attributes since repeating the steps leading from two to three attributes for going from three to four attributes most probably will result in partition equations which are still less solvable than those for, in total, three attributes. Moreover, the number sets to be split up will become smaller and smaller, ending up with trivial cases. The same argument holds, of course, for the basically simpler case of non-overlapping H_i. In addition, further considerations (not presented here) show that there are very restricting and complex conditions on choosing different secondary and/or tertiary attributes for different groups; so this case, too, will not be discussed here.

5. Conclusions

As – hopefully – demonstrated in the preceding sections, keys with NULL values in relations are possible using the notion of a maximal set of pairwise incompatible tuples each of which contains at least one NULL value and constructing such sets using the notion of attributes distinguishing between tuple sets - called "groups" - each of which is characterized by having NULL values for the same attributes.

It should be stressed, however, that the keys (with NULL values) considered here are keys of single Entities, in the sense of the Entity-Relationship-Model, or, alternatively, keys of base tables of a Relational Database. Composite keys, which are needed for representing many-to-many relationships, may contain NULL values only if the corresponding keys of their base tables already contain NULL values. Thus, for the case of one-attribute keys so common in real Relational Databases for the base tables the "entity integrity rule" mentioned in the Introduction, of course, holds, i.e. no NULL values are allowed for these keys and, correspondingly, the composite keys, too, do not contain NULL values. If, on the other hand, e.g. semantical considerations should lead to NULL values in the keys of base tables the considerations presented here will show how to handle them.