The strange case of The Pigeon-hole Principle

If you come to think about it, The Pigeon-hole Principle presents a case that is curious in more than one respect.

It has a remarkably high status: it has a special name, under which it is well-known; Ross Honsberger even calls it "A fundamental tool of combinatorics".

It is surrounded by some mystique. Proofs using it are often regarded as something special, something particularly ingeneous. This feeling of awe is, for instance, nicely reflected by Honsberger when he concludes one of his examples with " (Another triumph of the pigeonhole principle.)".

Its usual formulation is −as we shall see in a moment− unfortunate. Let me quote Honsberger once more − there is nothing special about me quoting Honsberger all the time: one of his books just happens to lie on my desk− :

> "if more than n objects are distributed into a set of n compartments, some compartment must receive [for "must receive" read "receives", EWD] more than one of the objects" .

(Instead of "more than n" one often encounters the more specific "n+1".)

My feeling is that these three observations are

not unrelated.

<div align="center">

*   *

*

</div>

To begin with, let me formulate the Purified Pigeon-hole Principle (PPP):

"For a non-empty, finite bag of numbers, the maximum value is at least the average value."

It is a generalization, because in the old formulation, it would be (OPPP)

"If more than n·k objects are distributed into a set of n compartments, some compartment receives more than k of the objects."

I think it is illuminating to analyze the differences between those two formulations (in what I consider increasing order of seriousness:

● PPP does not mention n and k explicitly; this eases the application, because the conscientious application of OPPP would require explicit instantiations of n and k

● OPPP is very operational; it would have been less so, had "are distributed" and "receives" had been replaced by "have been distributed" and "contains" respectively. The metaphor drags a process into the picture, where only the final state matters. (This is really serious, for what if "the objects are already in the com-

partments"? Well, of course we could have put them there, but then some people feel obliged to argue that we could have done so in "arbitrary" order: irrelevantia in the formulation of a theorem have the awkward consequence that the conscientious user of the theorem has to demonstrate their irrelevance.)

- OFPP mentions "objects" and "compartments", things that have nothing to do with the theorem. I know that some people believe that the introduction of such things "ease the visuali- sation", but I think the price too heavy to pay. When dealing with natural numbers, we should be able to deal with them in their own right and not be forced to introduce "objects" counted by them! (It is really like representing natural numbers in a unary number system, with the well-known difficulty of "the invisible zero".) Also, one may have to invent "compartments" that don't "look" like compartments at all. The "mystique" of the Pigeon-hole Principle is partly due the sometimes surprising inven- tion of the "objects" and the "compartments".

*   *

*

Let me now relate to you two stories about the Pigeon-hole Principle that both amazed me; for me — but perhaps I am sensitive to the point

of allergy — they suggest that the traditional Pigeon-hole Principle is a focal point of how mathematics is (but should not be) taught.

A few weeks ago I was told "The Problem of the Soccer Toto". (Not knowing the rules of the soccer toto, I got the explanation; for your bene-fit I shall repeat them.)

On a given Sunday, 13 soccer matches will be played. Each match has one of three possible outcomes, denoted by "1", "2", and "X" respectively. Hence there are 3^{13} possible Sundays. For our purpose "Sunday" will be one of those 3^{13} possible "columns" of height 13, with in each line one of the three permissible entries 1, 2, or X.

A toto form consists of a row of blanc columns of height 13; its user is free in his choice of the number of columns he is going to fill in. The question is to devise a strategy for filling in the minimal number of columns so that at least 1 column coincides with Sunday in at least 5 entries. Here we go.

Forget about the 5: what if at least 1 column should coincide with Sunday in at least 1 entry? Two columns won't do, for Sunday then could contain for each line an outcome missing in the corresponding line of the toto form. With 3 columns such that at least one row contains all three outcomes, at least one coincidence — in that row —

4

is guaranteed. So far we needed 3 columns and have used only one row. Can we use the remaining 12 rows to increase the score? By filling those each with some permutation of the three outcomes, we force in each row precisely one coincidence with Sunday. Make in each column the count of coincidences: their sum being 13, their average is $4\frac{1}{3}$ and their maximum (PPP) is ≥ 5. End of argument (heuristics included). Notice that their was no need at all to interpret the form as a table defining how, indeed, each object (= match) was placed in one compartment (= column).

But suppose that the question had been different: given a form in which 4 columns have been filled in such that each row contains each of the three possible outcomes, can you show that at least 1 column coincides in at least 4 entries with Sunday? Now the total number of coincidences is ≥ 13, the average count is $\geq 3\frac{1}{4}$, and hence the maximum count is ≥ 4. Please notice that in this case the metaphor of putting objects (= matches) into compartments (= columns) creates problems: an object may "occupy" two compartments! (It is utterly predictable how the average mathematician will save his face: "Sure! If you generalize the problem. I don't take the matches but the coincidences as my objects! I would have done so right at the start, would I

have known that you would generalize the problem.")
So much for the avoidable obligation of identifying
the "objects".

But now the sad part of the story. The colleague
that posed me this problem had been challenged
with it by one of his graduate students. When
we met again he was utterly amazed to hear that
the problem could be solved without pen and
paper and when I told him my solution, he looked
so puzzled that I am not sure he understood it.
"Of course I know the Pigeon-hole Principle, but
I never .....", and then wandering thoughts prevent-
ed him from completing the sentence. Was he
still looking for the objects and the compartments?
The way he had been introduced to the principle
and all the imaginations that went with it was
obviously was obviously a barrier to its straight-
forward application. I felt that I had had a glimpse
of something frightening.

The second story concerns my preparation of a
lecture on the Pigeon-hole Principle: I wanted to
show my students the most spectacular applica-
tion of it I had ever encountered. I scratched
my memory, and then I remembered: the lower
bound for the length of the longest monotonic
subsequence! I remembered my thrill, but had
forgotten the argument, which I then set out to
construct.

Let me state the problem first. We consider a sequence of numbers a.i with $0 \le i < N$. We get a subsequence of length n by removing some N-n elements from the sequence and retaining the remaining ones in their original order. It is called "an upsequence" provided for any a.i and a.j in the subsequence we have

$$i < j \Rightarrow a.i \le a.j \qquad ;$$

in the case that we have

$$i < j \Rightarrow a.i \ge a.j \qquad ,$$

it is called "a downsequence".

Let U be the length of a longest upsequence contained in the given sequence, and let D be the length of a longest downsequence contained in it. Prove that $N \le U \cdot D$ holds.

The argument I came up with went as follows. Let us construct $U \cdot D$ different labels; if we can now devise a regime that assigns to each element a distinct label, then N —being the number of labels used— is at most $U \cdot D$ —being the number of labels available.

How do we construct $U \cdot D$ different labels? The simplest way I can think of is all the integer pairs (u, d) with $1 \le u \le U$ and $1 \le d \le D$. How do we devise a regime that assigns for $i < j$ different labels to a.i and a.j?

7

In order to relate this regime to up- and down-sequences, we observe

$$i < j \Rightarrow a.i \leq a.j \vee a.i \geq a.j$$

i.e. $a.j$ can be used to extend either an upsequence or a downsequence ending at $a.i$. This observation reveals the regime: assign to $a.i$ the pair (u, d) with u (and d) the maximum length of an upsequence (and a downsequence respectively) that ends at $a.i$. This guarantees
(i) that distinct elements get distinct labels
(ii) that each label (u, d) needed satisfies $1 \leq u \leq U$ and $1 \leq d \leq D$.

I was quite pleased with the reconstruction of this argument, until I realized that what used to be "a triumph of the Pigeon-hole Principle" no longer used the Pigeon-hole Principle at all!

I could reintroduce an appeal to the Pigeon-hole Principle, but only by a contorted rephrasing. [Identify the elements with objects, the labels with compartments; assume N — the number of objects — to exceed $U \cdot D$ — the number of compartments —; then — PP — at least one compartment would contain more than one objects, which contradicts that distinct elements get distinct labels. Hence N does not exceed $U \cdot D$.]

Remark It is in this connection noteworthy that no one I asked formulated the Pigeon-hole Principle

as follows: "With objects distributed into compartments such that each compartment contains at most one object, the number of objects is at most the number of compartments". It is logically equivalent to the original formulation, but looks quite different. And that, presumably, is precisely the trouble. (End of Remark.)

With its physical, object-oriented formulation, the classical Pigeon-hole Principle is very vivid, almost catchy. And there lie precisely its major short-comings: the problem caused by such object-oriented formulation is that $A \Rightarrow B$ and the logically equivalent "counter-positive" $\neg B \Rightarrow \neg A$ invite completely different visualisations. The moral of the story seems to be that we should sharply distinguish between good mathematics and public relations.

Austin, 19 September 1986

prof. dr. Edsger W. Dijkstra
Department of Computer Sciences
The University of Texas at Austin
Austin, TX 78750 - 1188
United States of America

9