On bags and identical twins

Identical twins present a problem of nomenclature. Even when you have selected two different names, how do you decide which to give to whom? Does it matter? If, at the age of 5, they would decide to exchange their given names, would that make a difference? Would you notice it? (Would they notice it?)

* * *

With dummies ranging over sets, we are fine. For instance,
with W of type: set of ordered pairs of natural numbers
with w of type: ordered pairs of natural numbers
with p, q of type: natural number ,
the boolean expressions

$$\langle \forall w : w \in W : P.w \rangle$$

and

$$\langle \forall p,q : (p,q) \in W : P.(p,q) \rangle$$

are equivalent. (For a proof, see for instance EWD1064. At the time, I was a bit annoyed at the proof's length —about half a dozen steps— but in the mean time I am reconciled with it.)

The explicit statement of the types of W and ω, and, each time, of the range $\omega \in W$ tend to get abbreviated to something like "with W a set of ordered pairs of natural numbers and ω understood to be ranging over W", where the last clause is understood to settle the type of ω.

But with bags we have to be more careful — much more careful, as a matter of fact, than I (and others) have been in the past — . Under the analogous proviso "with $\mathcal{B}$ a bag of natural numbers and x and y ranging over $\mathcal{B}$",

$$x \in \mathcal{B} \quad , \quad y \in \mathcal{B}$$

are traditionally accepted as boolean expressions as legitimate as $\omega \in W$, and to express that each element in $\mathcal{B}$ is less than 100 I would not have hesitated to write

$$\langle \forall x : x \in \mathcal{B} : \; x < 100 \rangle$$

or, leaving the range under the proviso understood,

$$\langle \forall x :: \; x < 100 \rangle \quad .$$

But now the ground has become very slippery. The last expression formulates a constraint on $\mathcal{B}$; another constraint on

B would be that all its elements are the same:

$$\langle \forall x,y :: x = y \rangle \quad .$$

If we wish to express that any two distinct elements are equal, we can write with an explicit range the expression (by "trading" equivalent to the above)

$$\langle \forall x,y : x \neq y : x = y \rangle \quad .$$

But how slippery the ground has become, becomes clear if we wish to express that B contains no duplicates, i.e. that distinct elements in B are different. Don't try to do that by writing

$$\langle \forall x,y : x \neq y : x \neq y \rangle \quad ,$$

for this expression is equivalent to true, no matter what we choose for B .

The moral of the story is: <u>don't quantify over a bag</u>. We can quantify over a set, but with a bag two sets can be associated with, in general, different cardinality: there is the set of elements and the (possibly) smaller set of their values. In the previous paragraph we referred to expressing "that distinct elements in B are different", a fishy sentence to say the least. I would be much happier with the constraint in question being formulated by the requirement "that different elements have different values"; here the two nouns — "elements" and "values" — denote the two sets.

In EWD1103 "For the record: ETAC and the couples", we did the right thing. The problem is as follows. Of a finite number of couples we are told the following two things
(i) the oldest man is as old as the oldest woman
(ii) if two couples swap wives, the youngest of the one new couple is as young as the youngest of the other new couple.
Show that in each couple, man and woman have the same age.

To all intents and purposes, each couple is and ordered pair of two ages, but our <u>set</u> of couples corresponds in general to a <u>bag</u> of age pairs. ETAC did the right thing, continuing with "Let dummies x,y range over the couples; let the ages of man and woman of couple x be denoted by m.x and f.x respectively. Our proof obligation is then

$$\langle \forall x :: \ m.x = f.x \rangle \quad . \text{''}$$

As an example of a somewhat less fortunate formulation I quote the first pages of Chapter 4 from A.J.M. van Gasteren's Ph.D. Thesis (which I approved).

4 In adherence to symmetry

This chapter is another illustration of the complications engendered by the introduction of nomenclature, here emerging in the form of overspecificity and loss of symmetry. It also discusses the choice between recursion and complete unfolding.

We consider couplings, i.e. one-to-one correspondences, between two equally sized finite bags of natural numbers. Hence, a coupling can be considered a bag of —ordered— pairs of numbers, the subbags of which are as usual called its subcouplings. The value of a coupling is defined recursively by

- –the value of an empty coupling is 0;
- –the value of a one-element coupling is the product of the members in the single pair;
- –the value of a non-empty coupling is the value of one element $+$ the value of the remaining subcoupling.

Note . By the associativity and symmetry of addition, the above is a valid definition.
End Note .

The problem is to construct a coupling with maximum value. Such a maximum value exists, since the finite bags have a finite number of couplings.

A construction follows from the two lemmata

Lemma0 . Each subcoupling of a maximum coupling is itself a maximum coupling.
Lemma1 . In a maximum non-empty coupling, the maximum values of the two bags form a pair.

The construction then consists in choosing the maximum values to form a pair and constructing a maximum coupling for the remainders of the bags in the same way. The construction terminates since the bags are finite and decrease in size at each step.

Proof of Lemma0 . By the symmetry and associativity of addition we have —with $\cup$ for bag union—

the value of coupling $B \cup C$ = the value of B + the value of C;

Lemma0 now follows from the monotonicity of addition.

Proof of Lemma1 . We consider a maximum coupling, in which the maximum values U and V of the bags being coupled are paired with v and u respectively, and we prove $v = V \vee u = U$.

- If (U, v) and (u, V) are the same element of the coupling, $U = u \wedge v = V$.
- If (U, v) and (u, V) together form a two-element subcoupling we have

$$
\begin{array}{ll}
 & \textbf{true} \\
= & \quad \{\text{by Lemma0 and maximality of the coupling}\} \\
 & \text{value of } \{(U,v)\,,(u,V)\} \;\; \geq \;\; \text{value of } \{(U,V)\,,(u,v)\} \\
= & \quad \{\text{definition of "value"}\} \\
 & U*v+u*V \geq U*V+u*v \\
= & \quad \{\ \} \\
 & (U-u)*(v-V) \geq 0 \\
= & \quad \{(U-u)*(v-V) \leq 0 \text{, since } U \geq u \ \wedge \ V \geq v\} \\
 & (U-u)*(v-V) = 0 \\
= & \quad \{\ \} \\
 & U = u \ \vee \ v = V \quad .
\end{array}
$$

End Proofs .

* * *

What I am now not too happy about are the clauses "If (U,v) and (u,V) are the same element" and "If (U,v) and (u,V) together form ... ", as, linguistically, (a template of) the value is used to identify the element. If the original bags are {3,3} and {2,2} , the maximum coupling is the bag {(3,2), (3,2)}, (U,v) = (3,2) and (u,V) = (3,2) : are (U,v) and (u,V) "the same element"? I am afraid that I don't accept "It doesn't matter." for an answer.

With x and y ranging over the elements of a coupling, the value of element x being the ordered pair $(u.x, v.x)$, the fundamental theorem states that

$$\text{"the coupling is maximal"} \equiv \langle \forall x,y : u.x * v.x + u.y * v.y \geq u.x * v.y + u.y * v.x \rangle \quad ,$$

and, somehow, that seems to have been missed. The problem with prose is that its quantifications are so implicit.

Austin, 26 February 1993

prof. dr. Edsger W. Dijkstra
Department of Computer Sciences
The University of Texas at Austin
Austin, TX 78712 - 1188
USA