The bijection principle on the teaching of combinatorics

Michel Spira

April 28, 2008

Introduction

I have been teaching basic combinatorics for prospective middle and high school teachers for about twenty years now. Usually, these students have no training whatsoever in combinatorial thinking, having been taught that counting consists mainly of straightforward computations involving the multiplicative principle and *a priori* decisions as to which formula to use for a given problem – hence the unfortunately very common questions *Is this a combinations, arrangements or permutations problem?* and *Does order matter?*

These questions highlight one of the main problems students face in combinatorics: what is the type[1] of the objects one wants to count? In some situations this is simple; in a problem involving blouses and skirts, the ways Mary can dress herself are obviously in one to one correspondence with 2-sequences[2], the first position being a blouse and the second a skirt. One than counts such sequences and the problem is over. But, in general, combinatorics problems do not yield to such simple reasonings.

However, without one even noticing it, a powerful tool related to this problem has already made its appearance in the body of the previous paragraph. Let's spell it out clearly: *one does not count the set of ways Mary can dress herself; instead, one establishes a bijection between this set and a set*

[1]By *type* I mean standard elementary objects of combinatorics, such as sequences or subsets.

[2]I use *k-sequence* to mean *sequence of k symbols* or *of lenght k*.

of sequences, and then this last set is counted[3]. In other words, the objects to be counted in this problem are 2-sequences, and to do so standard tools of combinatorics (in this case, the multiplicative principle) are available.

This idea, usually refered to as the *bijection principle (BP)*, is used so often without explicit mention that, most of the time (say, as in the above problem of blouses and skirts), one is not aware of it. But, as everyone who has taught combinatorics should know, it is a tool of the utmost importance. I strongly believe that the BP should be stated and discussed in every single combinatorics textbook, regardless of the intended audience. I have read and worked with many such books but, for a long time, had seen the BP given its proper place only in Daniel Cohen's beatiful book *Basic techniques of combinatorial analysis* [1]. More recently, I was thrilled with A. T. Benjamim and J. Quinn's wonderful book *The art of combinatorial proof: proofs that really count* [2], in which it is clearly shown that the use of the BP can indeed become a form of high art. Hopefully other texts will follow in the same vein.

This article is written with the aim of describing a few ideas related to the BP which I found useful in actual teaching. I must say that nothing in here is new or original (except, maybe, the terminology "names"); all should be known to whoever has taught (or will teach) basic combinatorics. But, as I said before, having never seen things spelled out explicitly, I decided to try to do so.

I have done no research on the effectiveness of these ideas, nor have I developed them in any sort of theoretical framework. To be perfectly frank, I believe this to be unnecessary, since I cannot imagine that combinatorics can be taught at all without at least implicit use of this concept. All I can say is that over the years I have had the constant and strong feedback of my students, to the effect that they (the ideas) have changed - for the better! - their (the students') way of thinking about and of teaching combinatorics. I hope those reading this article will experience the same.

Last but not least, let me note that I am not advertising a one-way ticket to teaching or learning combinatorics. As will be clear from what follows, the BP is only one of a number of essential skills the student should have in order to become proficient in elementary combinatorics. The BP only works when paired with a solid grasp of the additive and multiplicative principles, a few well honed counting skills (say counting of sequences and subsets) and

[3]The reader will forgive me the use of *to count a set* instead of *to count the elements of a set*.

a repertoire of standard problems under one's belt. Since I do not want to talk about the teaching of combinatorics as a whole, I will stop here and get in business.

1 The bijection principle

Just to be formal, I will now state the BP.

The bijection principle (BP) *If there is a bijection between two sets then they have the same number of elements.*

I always illustrate the BP by putting the fingers of my hands tip to tip and asking what is the meaning of this gesture, to which my students almost invariably answer *You have five fingers in each hand.* I expect this wrong answer[4], which gives me the opportunity to clarify things and easily illustrate non-injective and non-surjective functions. Of course there is always a student who makes the (also expected) silly joke about Brazil's president missing a finger in one of his hands, but I quickly restore order and move ahead.

The BP is so much in the nature of a truism that the students' reaction is an immediate "So what?". The answer to this natural question is the philosophy behind the BP, which can be stated as follows. One is faced with the task of counting a set A, but for whatever reason this is difficult. "Luckily" (see below), it is possible to establish a bijection between A and another set B, and we know how to count B. Then A is automatically counted and we are done.

So far so obvious, so what is all the fuss about? It is now that one talks about the hard facts of life. First, it should be stressed that the BP is *not* a counting tool; as it is clear from its statement, it only changes the problem of counting A into the one of counting B. Secondly, finding B and a bijection $f : A \to B$ may be (and most of the time, is) a far from easy task. More than that, one does not find B first and f afterwards, or vice-versa; there is a symbiotic relationship between them and they are, of necessity, created simultaneously. And then there is the problem of proving that f is indeed a bijection.

[4]The right answer, of course, is that the right and left hands have the same number of fingers.

I usually begin to address these points with the following example, copied more or less *verbatim* from [1].

Example 1 *1500 teams compete in the soccer tournament of the Andromeda galaxy. The organizers let the teams know that every game must have a winner and that the team that loses a game is immediately excluded from the tournament. How many games will be played till the champion is known?*

Usually the students will ask about pairing rules, on the basis of reasonings such as "In the first round there will be 750 games, in the second 375 and then how does one make pairings for an odd number of teams?". There is general disbelief when they are told that this is irrelevant. Eventually, if need be with some strong hints, the idea of bijecting the set of games with the set of losers by *game* $\leftrightarrow$ *loser* eventually comes out, giving 1499 as the answer[5]. It takes some doing to convince the students that we have indeed a bijection, but eventually they yield, albeit with some suspicion; after all, this is supposed to be a difficult problem and it is unfair to solve it in such a simple way. But after a while the idea sinks in, and it should; this is indeed a striking application of the BP, in particular because of the utmost simplicity in counting the set of losers and the fact that, indeed, pairing rules are completely irrelevant.

Let's fix some terminology now. One wants to count a set of *objects* A by establishing a bijection f between A and a set of *names* B[6]; in order to count A one is then reduced to *counting names*. I also use *to every object corresponds a unique name* to mean *well defined* and after that *every name comes from a unique object* gives us *bijective* right away. This helps to keep the discussion about the bijectivity of f on a coloquial level, free of technicalities. As a matter of fact, when one gets a good choice of $f : A \to B$, the bijectivity is almost always self-evident and deserves only a cursory mention; it is usually besides the point and distracting to ask for formal proofs.

One now is led to the crux of the matter, which is the choice of names.

[5]An added bonus here is that this reasoning (and, in fact, all others in this article) works for n teams for any n, thereby illustrating another important principle in combinatorics, which I call the *generalization principle*. It can be loosely stated as *good combinatorial reasonings are those which hold when numbers are replaced by letters*. But this is the theme of another article.

[6]One could also use, say, *descriptions* instead of *names* and substitute *naming* by *describing*; other alternatives are possible and reasonable. To each his/her own choice of terminology.

2 The choice of names

What's in a name? For our purposes, a name is a way of describing an object within an agreed upon code of communication, respecting the principles of *precision* and *economy*; of course, names should also be amenable to (hopefully easy) counting. Loosely speaking, in assigning names the following conditions should be (ideally) respected.

First and obviously, the set of objects to be counted should be in one to one correspondence with the set of names. This is what is meant by precision: an object has a unique name and every name uniquely identifies the corresponding object. Secondly, all names should be of the same standard combinatorial type, so that a single counting technique is enough to count them all; I call this *uniformization*. And thirdly, names must be economical, in the sense that they should not contain redundant information. I guess there is no need to justify this requirement by arguments other than aesthetics alone, but if asked to do so I would say that redundant information may (and usually does) obscure the counting process and also conflicts (*ibid*) with uniformization (cf. example 6 below).

Precision and economy are, in a way, opposite to each other. One is always tempted to "make sure" by adding information at the expense of economy, or else to be so economical in giving information that precision is compromised. The (unique) middle point between these concepts is where good names come in. Putting it succintly, good names are those that contain only necessary and sufficient information for recovering their corresponding objects.

In order to make the students add all this to their thinking in a practical way, I ask them find a way to describe to a friend, over the telephone, one or more of the objects to be counted, keeping in mind that telephone rates are very high. This works: if your friend does not know which object you are talking about, you are being imprecise and you have to repeat or to clarify, losing money in the process. Likewise, if you send redundant information, you are wasting time and you lose money too.

Now for some examples. For them I have chosen to present solutions which illustrate the use of the BP; others, of course, are possible [7]. These examples look simple, but I have found all of them to be consistently difficult

[7] I have also chosen to use small numbers for simplicity, but it should be clear that all reasonings work in general .

to beginning students, even for those who are sufficiently conversant with basic principles. After following these examples, the students always get the hang of the "naming" idea, which can then be relegated to the unconscious and called forth when needed.

I note that the pace of the examples below is not the way I work them out in the classroom. There one discusses at lenght, waiting and commenting on the student's ideas and eventually, when things come to a standstill, giving partial ideas so as to let them proceed further. So the solutions I give should be read as rough sketches of discussions to be developed in class.

Example 2 *How many 5-sequences can be made with the digits* $1, 1, 1, 8, 9$*?*

In order to describe such sequences we first discuss economical ways to name them. After a while, the idea of first saying in which position (say, from the left) is the 8 and then do the same for the 9 emerges; for instance, $19118 \leftrightarrow 52$ and $89111 \leftrightarrow 12$. In order to discuss the bijection part, it is enough to ask something like "Which sequence corresponds to 42?" and wait for the right answer. Note that precision and non-redundancy are obvious here, since once we know where the 8 ad the 9 are, there is no need to say anything about the position of the 1's. In this way we establish a bijection between the set of our sequences and the set of 2-sequences of distinct symbols chosen among $\{1, 2, 3, 4, 5\}$; this last is easy to count, and we are done.

Example 3 (circular permutations) In how many ways can 9 people sit around a circular table, all seats being identical?

Here there is always the problem of the meaning of *ways of seating*, which is why this problem is always stated with the no more illuminating restriction of "up to rotation". What I try to do to clarify things is to make the students imagine themselves standing on the center of the table, turning round – always clockwise, say – and pointing out, in order, the the people they see. In this way, all realize that a way of seating can be described by a 9-sequence of people. But then, maybe after a 10-second delay, the whole class protests that a given way of seating can correspond to many such sequences; in other words, they say that an object can have more than one name, which is bad – and one of the main points of this exercise has already been made, with no effort whatsoever. How can we get rid of this bad choice?

Two strategies then offer themselves. The first depends on sticking to what we have and discovering how many possible names a given way of seating can have; I will defer that to the next section. The second does things by a neat trick, which is not so much of a a trick when seen from the point of view of uniformization. One numbers the people from 1 to 9 and then uses only names (i.e., 9-sequences) which begin with 1. Ideally (or with some prodding, if need be) someone will then raise his/her hand and invoke non-redundancy to argue that the beginning 1 is redundant. This finishes the discussion: we are counting permutations of the symbols 2, 3, . . . , 9 and the game is over.

Example 4 *In how many ways can 10 identical candies be distributed among 4 (obviously distinct!) children?*

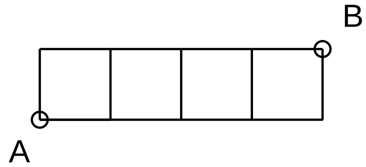
This is a standard combinatorial problem, which for the beginning student seems to defy solution using only basic tools. And yet, a strikingly beautiful use of the BP makes things surprisingly simple. Let's first order the children and fix this order for the rest of the problem. Say then the children got 3, 4, 1 and 2 candies, in this order; we name this as OOO+OOOO+O+OO. If the children got 4, 1, 0 and 5 candies, we would write OOOO+O++OOOOO. One should now offer a couple of sequences of O's and +'s and ask which candy distributions they came from, which will readily convince the students that we have indeed a bijection. At the end, we are counting the number of 10-sequences made up of ten O's and three +'s, and this is a standard problem, the answer being $\binom{10+3}{3}$.

It should be noted that in this example the names offer themselves to the careful student. It is enough to think of telling a friend about a given candy distribution among the children with as few words as possible; after having agreed on a fixed ordering of the children, one would say "three (pause), four (pause), one (pause), two" which, of course, *is* OOO+OOOO+O+OO.

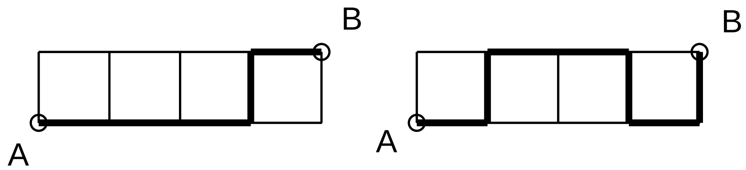
Example 5 *In how many ways can 10 different rings be distributed among the fingers of one hand? It is assumed that any finger can hold all rings.*

Let's be quick here, since this example is a variation of the previous one. One reads aloud the the 10-sequence of rings, from thumb to little finger and from the base of the fingers to their tips. This is still not a good name, but in the act of reading the jumps between fingers are marked by pauses, and the right names offer themselves again, in this case being 14-sequences of ten distinct symbols (the rings) and four identical pauses. This is standard counting again, the answer being $\frac{14!}{4!}$.

Example 6 *In how many ways can one go from point A to point B, walking along the segments of the figure without going twice over the same segment?*



This is a prototypical problem in combinatorics and specially well suited for illustrating the BP. In the picture below we show two ways of going from A to B. It is natural to describe a path using R for right, U for up and D for down; the first path is then RRRRUR and the second RURRDRU.



However, there is a problem here: our names are not of the same combinatorial type, since they do not have a fixed length. No uniformization and counting is out of question. It takes some time to see that the problem here is redundancy; indeed, after U or D, only R can follow, so the subsequences UR and DR are already redundant. One tries DR$\mapsto$D and UR$\mapsto$U, but it is clear that this doesn't help. This is, indeed, a tough problem, and a great one to work with because the students usually get quite excited about it – it just looks *so* easy!

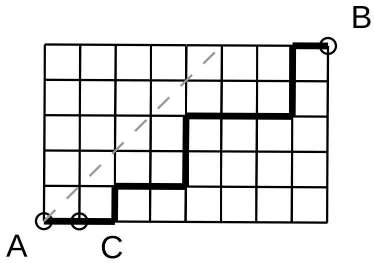
Eventually (again, maybe with some prodding) one hits the right idea: it suffices to say, for each square, if one crosses it walking on the bottom or on the top, so that the first path would be named BBBT and the second BTTB; our names are then 4-sequences of B's and T's. One takes care of the bijection part by a few examples and the work is over. This problem and the order of ideas I sketched above has never failed to make a profound impression upon the students.

The last example of this section is the famous *ballot problem*. It presents two beautiful application of the BP, both being rich sources of ideas in combinatorics.

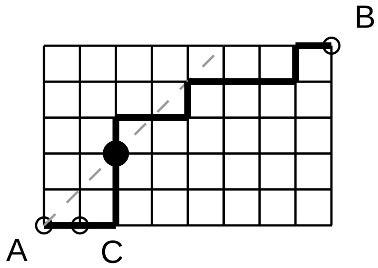
Example 7 *A class of $m+n$ students, with $m > n$, made a poll to elect its class president. There were two candidates, Mary and Nelville, and when all votes are counted Mary had been chosen with m votes against n for Nelville.*

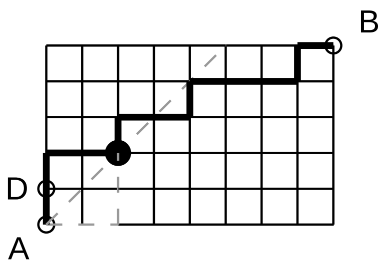
In how many ways could the the votes having be counted, one by one, so that Mary would always have been ahead of Nelville?

One first uses the BP to see a ballot as a nondecreasing path (i.e., a path which goes only right and up) from A to B on an $m \times n$ grid, votes for Mary being represented by horizontal segments and those for Nelville as vertical ones. Say $m = 8$ and $n = 5$ for convenience; in the 8×5 grid below the path shown represents a ballot in which the first two votes were Mary's, the next one Nelville's, the next two Mary's and so on. Ballots in which Mary was always ahead of Nelville correspond to paths from C to B which do not touch or cross the dotted diagonal.



Say now we have a bad path; it begins in C and then touches or crosses the dotted diagonal in one or more points. We select the first such and then reflect the portion of the path up to this point on the diagonal, leaving the remaining part untouched, as illustrated below; the first bad point, on which the reflection is based, is shaded. Note that every path obtained in this way passes through the point D.





In this way we biject the set of bad paths with the set of paths from D to B; this beautiful application of the BP is known as *the reflection principle*. I should remark that it takes some doing to get everyone happy about the fact that we have, indeed, a bijection. To finish we use the *good* $=$ *all* $-$ *bad* version of the additive principle; standard counting then yields the answer

$$\binom{m+n-1}{m-1} - \binom{m+n-1}{n-1} = \frac{m-n}{m+n}\binom{m+n}{n} .$$

3 Many-to-one and one-to-many correspondences

In this short section I want to discuss an immediate and very useful generalization of the BP and a few of its applications.

It is quite common to discuss example 1 of the previous section as follows: "Think of the 1's as being different, say painted in different colors. Then one has five different objects, which can be permuted in 5! different ways. Now since the 1's were originally identical, one must divide by 3! to account for the error introduced by thinking they were different, and we get the answer $\frac{5!}{3!}$". Faced with such *ad hoc* reasoning, all that remains for the student is to accept the argument on faith and proceed in similar problems by analogy, but without real understanding and the corresponding lack of creativity in not-so-analogous situations.

At any rate, this is a good general technique in combinatorics: to solve a related problem, not quite the one we want, and then get the answer to ours introducing, somehow, a "correcting factor". The point is to find a convincing way to explain how and why those factors appear and why they are used in the way they are. In this section I want to present some examples in which one divides by a correcting factor, based on a simple generalization of the BP. All we need to do is relax a little and stop being so picky about our functions being bijections or, for that matter, being functions!

Let's think about a surjective function $f : A \to B$ such that all elements of B have inverse images of the same cardinality k; I call this a *k to 1 correspondence*. It is then clear that $|A| = k|B|$[8]. This can be made clear to the students by some illustrative examples with small k (in my experience, simple drawings of functions with arrows are more than enough); no need for a formal proof. The same situation with arrows reversed and A and B interchanged gives us a "surjective correspondence" $f : A \to B$ in which every element of A has exactly k images and image sets of different elements are disjoint; of course I call this a *1 to k correspondence*. In this case we get

[8] I denote by $|S|$ the cardinality of a set S.

$|A| = \frac{|B|}{k}$[9]. It is convenient to refer to those two situations as *homogeneous.*

Now for some examples, including some of the previous section, but now treated according to the new theme. As before, I will present bare sketches of the way I work them in class.

Example 8 *How many 5-sequences can be made with the digits* $1, 1, 1, 8, 9$*?*

One begins with the set A of all permutations of $1_{red}, 1_{blue}, 1_{green}, 3, 5$ and then gives each permutation a name by painting all the 1's white. This was bad before, many objects sharing the same name; but homogeneity is obvious and we now have a 3! to 1 correspondence, which is good. In particular, the answer $\frac{5!}{3!}$ is now seen clearly as arising from a general context.

Of course this is immediately generalized to get the formula for permutations of a multiset; one should not lose momentum.

Example 9 *In how many ways can 9 people sit around a circular table, all seats being identical?*

The work of example 3 showed us that each way of seating could be named in exactly 9 ways, the set of names being the set of permutations of the symbols 1, 2, ..., 9. Here we have a 1 to 9 correspondence, so that our answer is $\frac{9!}{9}$.

Example 10 *How many k-subsets*[10] *does a n-set have?*

This comes after one knows how to count sequences and having agreed to use A_k^n (not the formula!) as shorthand for the number of k-sequences of distinct symbols chosen from an n-set. A given k-subset can then have $k!$ names, obtained by "adding order" to its elements. In this way we have a 1 to $k!$ correspondence; the set of names having A_k^n elements, our answer is $\frac{A_k^n}{k!}$.

Alternatively, one can take the set of all k-sequences of distinct symbols chosen among the n available ones and then "take the order out". In this way the set of names becomes the set of the subsets we want to count. We now have a $k!$ to 1 correspondence and the answer follows as before. This, of course, is the easiest way to get the usual formula for $\binom{n}{k}$.

[9] It should be noted that the BP is a particular case of both these statements in the case $k = 1$.

[10] I use *k-(sub)set* to mean (sub)set of k elements.

References

[1] Daniel I. A. Cohen: **Basic techniques of combinatorial theory**. Wiley, 1978

[2] Arthur T. Benjamin and Jennifer J. Quinn: **Proofs that really count: the art of combinatorial proof**. Dolciani Mathematical Expositions, MAA, 2003.