A BIJECTIVE PROOF OF AN IDENTITY EXTENDING A CLASSIC RESULT OF HAJÓS

MIKLÓS BÓNA

Abstract. We provide bijective proofs of two classic identities that are very simple to prove using generating functions, but surprisingly difficult to prove combinatorially. The problem of finding a bijective proof for the first identity was first raised in the 1930s. The second, more involved identity takes the first one a step further.

Dedicated to the memory of Herb Wilf

1. Introduction

1.1. A Classic Problem. The obvious identity

\[
\frac{1}{1-4x} = \frac{1}{\sqrt{1-4x}} \cdot \frac{1}{\sqrt{1-4x}}
\]

implies, after equating coefficients of \(x^n\) on both sides, the identity

\[
4^n = \sum_{i=0}^{n} \binom{2i}{i} \binom{2(n-i)}{n-i}.
\]

Finding a bijective proof for (1) is surprisingly difficult.

Note that (1) has the following simple interpretation. There are as many ways for two teams to score a total of \(2n\) goals in a soccer game as there are ways for those two teams to reach a tie score at the end of the first half of the soccer game, then to hold an intermission, and then to reach an \(n-n\) tie at the end of the game.

Richard Stanley ([2], page 52) provides a brief history of the quest for a combinatorial proof of (1), mentioning that the problem was raised by Pál Veress and solved by György Hajós in the 1930s. Stanley also mentions a more recent combinatorial proof by Daniel Kleitman [1] and a discussion by Mártia Svéd [3].

As a “warmup” to our main result, in this paper we will provide a combinatorial proof of (1), by constructing a simple bijection. This result, and its proof, while similar in nature to our main result, are not necessary for understanding that main result. Readers so inclined may skip Section 2.

1.2. Our Main Result: An Identity One Step Further. The obvious identity

\[
\frac{1}{(1-4x)^{3/2}} = \frac{1}{\sqrt{1-4x}} \cdot \frac{1}{\sqrt{1-4x}} \cdot \frac{1}{\sqrt{1-4x}}
\]
implies, after equating coefficients of $x^n$ on both sides, the identity

$$ (2n + 1) \cdot \binom{2n}{n} = \sum_{i+j+k=n} \binom{2i}{i} \binom{2j}{j} \binom{2k}{k}, $$

where the sum on the right-hand side is taken over all triples of non-negative integers that sum to $n$.

Note that (2) has the following interpretation. There are as many ways for two soccer teams to play reach any score in the first half, then hold an intermission, and then reach an $n - n$ tie at the end of the game as there are ways for two ice hockey teams to reach a tie score at the end of the first period, hold an intermission, reach a tie score at the end of the second period, hold an intermission, and then reach a score of $n - n$ at the end of the game. (Soccer games have two halves, while ice hockey games have three periods. Two outcomes are considered identical if they agree in the score at the end of each half or period.)

In Section 3 we give a bijective proof of (2). The proof will have an interesting structure. After disposing of some trivial cases, we will construct two bijections, each of which will map half of the remainder of the domain onto half of the remaining range of the bijection that we intend to build. However, these two halves will not be disjoint. Fortunately, there will be a simple bijection between the intersection of these two halves and the complement of their union, which will enable us to complete our proof.

2. Warming Up. A Bijective Proof of the First Identity

In this section, we will work with northeastern lattice paths. A northeastern lattice path is a lattice path consisting of steps $(0, 1)$ (a north step) and $(1, 0)$ (an east step). The number of northeastern lattice paths from $(0, 0)$ to $(a, b)$ is clearly $\binom{a+b}{a}$ for all non-negative integers $a$ and $b$, whereas the number of northeastern lattice paths starting at $(0, 0)$ and consisting of $2m$ steps is $2^{2m} = 4^m$.

Therefore, identity (1) that we are going to prove in this section claims that the number of all northeastern lattice paths starting at $(0, 0)$ and consisting of $2n$ steps agrees with the number of northeastern lattice paths that start at $(0, 0)$, end in $(n, n)$, and are marked at some point $(i, i)$ on the main diagonal.

The observation we present in this paragraph was made by Kleitman in [1]. A northeastern lattice path $p$ starting at $(0, 0)$ and consisting of $2n$ steps has at least one point in common with the main diagonal $x = y$ (the point $(0, 0)$ is always such a point). Let us say that $N = (i, i)$ is the last (most northeastern) such point, and let $N$ split $p$ into parts $p_1$ and $p_2$. Then $p_2$ is a northeastern lattice path starting on the main diagonal $x = y$ that never touches the main diagonal again, while $p_1$ starts and ends in the main diagonal. Therefore, if we can find a bijection $b$ from the set of all northeastern lattice paths starting at $(i, i)$ and ending in $(n, n)$ to the set
of such \( p_2 \), then the map \( p \to (p_1, b^{-1}(p_2)) \) will be a bijection proving (1). Here \( (p_1, b^{-1}(p_2)) \) means the concatenation of \( p_1 \) and \( b^{-1}(p_2) \), marked at the point where these two paths meet.

We will construct such a bijection in the following theorem.

**Theorem 1.** There is a natural bijection \( F_n \) from the set \( X_n \) of lattice paths with \( 2n \) steps that start at \((0,0)\) and end in \((n,n)\) to the set \( Y_n \) of lattice paths that start at \((0,0)\), consist of \( 2n \) steps and never touch the \( x = y \) line other than in their starting point.

Note that the following interpretation of Theorem 1. There are as many ways for a soccer game to end in an \( n - n \) tie as there are ways for a soccer game to bring a total of \( 2n \) goals so that the score is never a tie other than at 0-0.

The following lemma is the main ingredient of our proof of Theorem 1.

**Lemma 1.** Let \( k \) and \( n \) be positive integers so that \( n \leq k \leq 2n \). Let \( A(n,k) \) be the set of all lattice paths from \((0,1)\) to \((k,2n-k)\). Let \( B(n,k) \) be the set of all lattice paths from \((0,1)\) to \((k,2n-k)\) that never touch the line \( x = y \).

Then there exists a natural bijection \( f_{n,k} : A(n,k) \to B(n,k) \cup A(n,k+1) \).

**Proof.** (of Lemma 1.) Let \( p \in A(n,k) \). If \( p \in B(n,k) \), then we set \( f_{n,k}(p) = p \). Otherwise, let \( P \) be the first point on \( p \) that is also on the line \( x = y \). Let us reflect the part of \( p \) that is between \((1,0)\) and \( P \) through the line \( x = y \), to get the path \( p' \). Then \( p' \) is a path from \((0,1)\) to \((k,n-k)\). Translate \( p' \) by the vector \((1,-1)\), to get the path \( p'' \) that is a path from \((0,1)\) to \((k+1,n-k-1)\). Then set \( f_{n,k}(p) = p'' \).

See Figure 2 for an illustration.

It is straightforward to see that \( f_{n,k} \) is a bijection. Indeed, each element \( q \) of \( A(n,k+1) \), when translated by the vector \((-1,1)\), will intersect the line \( x = y \), so the unique \( f_{n,k}^{-1}(q) \) can be easily recovered. \( \square \)

Now we are in a position to prove Theorem 1.

**Proof.** (of Theorem 1.) Let \( p \in X_n \), and let us assume without loss of generality that \( p \) starts with an east step. Keep the notation of the proof of Lemma 1. Then, identifying \( p \) with the path obtained from \( p \) by removing its predetermined first step, \( p \) is an element of \( A(n,n) \). So \( f_{n,n}(p) = p_1 \) is defined, and \( p_1 \in B(n,n) \cup A(n,n+1) \).

If \( p_1 \in B(n,n) \), then we set \( F_n(p) = p_1 \) and we stop. Otherwise, that is, if \( p_1 \in A(n,n+1) \), then we repeat this procedure. More precisely, we consider \( p_2 = f_{n,n+1}(p_1) \in B(n,n+1) \cup A(n,n+2) \). If \( p_2 \in B(n,n+1) \), then we set \( F_n(p) = p_2 \) and we stop, while if \( p_2 \in A(n,n+2) \), then we continue by considering \( p_3 = f_{n,n+2}(p_2) \in B(n,n+1) \cup A(n,n+2) \).

If we have not stopped by the \( i \)th step, then in the \( i \)th step, we consider \( p_i = f_{n,n+i-1}(p_{i-1}) \in B(n,n+i-1) \cup A(n,n+i) \). If \( p_i \in B(n,n+i-1) \), then we set \( F_n(p) = p_i \) and we stop. If \( p_i \in A(n,n+i) \), then we go to step \( i + 1 \), where we consider \( p_{i+1} = f_{n,n+i}(p_i) \) according to the same rules. The
procedure stops and \( F_n(p) \) gets defined when \( p_i \in B(n, n + i - 1) \), that is, when the current image of \( p \) stays below the main diagonal.

See Figure 2 for an illustration.

This procedure will define \( F_n(p) \) for all \( p \in X_n \) starting with an east step. Indeed, note that \( A(n, 2n) = B(n, 2n) \), so all paths \( A(n, 2n) \) are below the main diagonal. So in at most \( n \) steps, \( F_n(p) \in Y_n \) will be defined for all \( p \in X_n \) that start with an east step.

If \( p \) starts with a north step, simply reflect \( p \) through the line \( x = y \), to get \( p' \). Then proceed as above, to obtain \( F_n(p') \). Finally, reflect \( F_n(p') \)
through the line $x = y$ to get $F_n(p)$. Note that in this case, $F_n(p)$ will be entirely above the main diagonal, except for its starting point.

We will now show that $F_n$ is a bijection by showing that it has an inverse. Let $u \in Y_n$, and let us assume without loss of generality that $u$ starts with an east step. That means that $u$ ends below the main diagonal, in the point $(m, 2n-m)$ for some $m > n$. So $u \in B(n, m)$, and in particular, $u \in A(n, m)$. Therefore, the unique path $u_1 = f_{n,m-1}^{-1}(u) \in A(n, m-1)$ exists. In order to get $u_1$, translate $u$ by the vector $(-1, 1)$. Let $U$ be the first intersection point of the line $x = y$ and the translated copy of $u$. Then reflect the part of $u$ that is before $U$ through the line $x = y$ to get $u_1$. Then proceed in a similar manner. That is, since $u_1 \in A(n, m-1)$, the unique path $u_2 = f_{n,m-2}^{-1}(u) \in A(n, m-2)$ exists, then the unique path $u_3 = f_{n,m-3}^{-1}(u) \in A(n, m-3)$ exists, and so on. After $m-n$ steps, we get a path $u_{m-n} \in A(n, n)$, and that is the unique path satisfying $F_n(u_{m-n}) = u$. □

See Figure 3 for the algorithm that defines the map $F_n$.

**Figure 3.** The algorithm that produces $F_n(p)$. Note that the algorithm always stops after at most $n$ increases of $k$.

### 3. Our Main Result: A Bijective Proof of the Second Identity

The goal of this section is to give a bijective proof of formula (2).

The lattice paths we use in this section will have two allowed steps, $(1, 1)$ (an “up” step) and $(1, -1)$, a “down” step. For the rest of this paper, when we say “lattice path”, we mean a lattice path of that kind.

Let $T_n$ be the set of ordered triples of lattice paths $P = (A, B, C)$, where $A$ is a lattice path consisting of $i$ up steps and $i$ down steps, $B$ is a lattice path
consisting of $j$ up steps and $j$ down steps, and $C$ is a lattice path consisting of $k$ up steps and $k$ down steps, for some ordered triple $(i, j, k)$ of non-negative integers satisfying $i + j + k = n$. Clearly, $|T_n| = \sum_{i+j+k=n} \binom{2i}{i} \binom{2j}{j} \binom{2k}{k}$, which is the left-hand side of the identity that we intend to prove. We will place our triples $(A, B, C)$ so that $A$ starts at $(0, 0)$; then $C$ will end at $(2n, 0)$.

Let $D_n$ be the set of ordered pairs $(H, X)$, where $H$ is a lattice path consisting of $n$ up steps and $n$ down steps, and $X$ is one of the $(2n + 1)$ lattice points on $H$. It is then clear that $|D_n| = (2n + 1)\binom{2n}{n}$, which is the right-hand side of the identity to be proved. Again, we will place the elements of $D_n$ so that $H$ starts at $(0, 0)$ and ends at $(2n, 0)$.

We are going to define a bijection $g : T_n \rightarrow D_n$ in three steps, the “trivial step”, the “core step”, and the “correction step”.

3.1. The Trivial Step. Let $R_n \subseteq T_n$ denote the subset of $T_n$ that consists of pairs $(A, B, C)$ in which $C$ is empty. Let $D_n^0 \subseteq D_n$ be the subset of $D_n$ consisting of pairs $(H, X)$ in which $X$ is on the horizontal axis. For elements of $R_n$, we simply define $r(A, B, C) = r(A, B, \emptyset)$ to be the pair $(H, X) \in D_n^0$ in which $H$ is simply the concatenation of $A$ and $B$, and $X$ is the point at which $A$ and $B$ meet. The following proposition is then obvious.

**Proposition 1.** The map $r : R_n \rightarrow D_n^0$ defined in the previous paragraph is a bijection.

3.2. The Core Step. First, we define a map on a large subset (almost half) of $T_n$. Let $U_n$ be the subset of $T_n$ that consists of elements $(A, B, C) \in T_n$ so that (exactly) one of the following two conditions hold:

1. The path $B$ ends with a down step and the path $C$ has at least one point strictly above the horizontal axis, or
2. The path $B$ is empty, the path $C$ has at least one point strictly above the horizontal axis.

Let $D_n^+$ denote the subset of $D_n$ consisting of pairs $(H, X)$ in which $X$ is a point of $H$ that is strictly above the horizontal axis.

Let $s : U_n \rightarrow D_n^+$ be defined as follows. The *height* of a lattice point is simply its vertical coordinate. Let $(A, B, C) \in T_n$ and let us assume first that the first criterion is satisfied, that is, $B$ ends with a down step and $C$ has at least one point strictly above the horizontal axis. Let $K$ be the leftmost point on $C$ that has maximal height. Let this height be $k$, then we it follows from our assumptions that $k > 0$. Let $K$ cut $C$ into the left part $C_1$ and the right part $C_2$. Now we define $s(A, B, C) = (H, X)$, where $H$ is the concatenation of the paths $C_1$, $A$, $B$, and $C_2$ (each path starts at the endpoint of the preceding path), and where $X$ is the point where $A$ and $B$ meet. See Figure 4 for an illustration.

If $B$ is empty, then process is identical to the one above. Note that because $B$ is empty, $X$ will be the starting point of $C_2$. 
Note that in both cases, $X$ has the same height $k$ as $K$, and that $K$ is the leftmost point of $H$ on the horizontal line $y = k$. In the special case when $A$ is empty, the points $X$ and $K$ coincide.

**Example 1.** Figure 4 illustrates how the map $s$ is defined. We drew the parts of $H$ that are defined by $A$, $B$, and $C$, with the respective line styles of $A$, $B$, and $C$.

![Diagram](image)

**Figure 4.** The map $s : U_n \rightarrow D_n^+$. 

**Lemma 2.** The map $s : U_n \rightarrow D_n^+$ defined above is a bijection.

**Proof.** Let $(H, X) \in D_n^+$. We will show that $(H, X)$ has exactly one preimage under $s$. Draw a horizontal line $\ell$ across $X$. By the paragraph immediately preceding our lemma, the point $K$ can be recovered as the leftmost common point of $\ell$ and $H$. Therefore, if $s(A, B, C) = (H, X)$, then $A$ can be recovered as the (possibly empty) part of $H$ that is between $K$ and $X$. Then $C_1$ can be recovered as the part of $H$ that is between $(0,0)$ and $K$.

We still have to show that there is exactly one way in which we can split the part of $H$ that is between $X$ and $Z$ into $B$ and $C_2$. Note that it follows from our definitions that if such $B$ and $C_2$ exist, and $B$ is not empty, then $B$ ends with a down step and $C_2$ starts with a down step. Indeed, $C$ is split into two parts $C_1$ and $C_2$ at a point of maximal height, so $C_2$ never goes above its starting height. In particular, this means that if $B$ is non-empty, then the point $Z$ where $B$ and $C_2$ meet is the rightmost point of $\ell$ strictly on the right of $X$ that is crossed by $H$, (that is, $H$ reaches $Z$ by a down step and leaves it by another down step). So if such $Z$ exists, then we recover $C_2$ and $B$ as the parts of $H$ between $X$ and $Z$ and $Z$ and $(2n,0)$. If such $Z$ does not exist, then $B$ is empty, and $C_2$ is the part of $H$ that is between $X$ and $(2n,0)$.

$\square$
Example 2. Figure 5 illustrates how to recover $s^{-1}(H, X)$ if $(H, X) \in D_+^n$ is given.

Now let $V_n$ be the set of triples $(A, B, C) \in T_n$ that are reflected images of the elements of $U_n$ through the horizontal axis. That is, $(A, B, C) \in V_n$ if (exactly) one of the following conditions hold.

1. The path $B$ ends with an up step and the path $C$ has a point strictly below the horizontal axis, or
2. the path $B$ is empty and the path $C$ has a point strictly below the horizontal axis.

Now we define a map $t : V_n \to D_n^-$ which is essentially the reflected image of $s$ through the horizontal axis.

Let $(A, B, C) \in V_n$. Let $k < 0$ be the minimum height that any point of $C$ has, and let $K$ be the leftmost point of $C$ with that height. Let $K$ cut $C$ into the left part $C_1$ and the right part $C_2$. Define $z(A, B, C) = (H, X)$, where $H$ is the concatenation of the paths $C_1, A, B,$ and $C_2$ and where $X$ is the point where $A$ and $B$ meet.

Again, if $B$ is empty, then process is identical to the one above. Note that because $B$ is empty, $X$ will be the starting point of $C_2$.

Just as we saw it when we defined $s$, the point $X$ has the same height $k$ as $K$, and $K$ is the leftmost point of $H$ on the horizontal line $y = k$. In the special case when $A$ is empty, the points $X$ and $K$ coincide.

Lemma 3. The map $t : V_n \to D_n^-$ is a bijection.

Proof. This can be proved in a way analogous to the way in which Lemma 2 was proved. Alternatively, note that if $r_x$ denotes reflection through the
horizontal axis, then

\[ t = (r_x \circ s) \circ r_x, \]

and the proof is immediate since all three maps on the right-hand side are bijective. □

3.3. The Correction Step. Our work is not finished, since \( U_n \) and \( V_n \) are not disjoint, and also, \( U_n \cup V_n \neq T_n - R_n \). (Two halves cannot cover a whole if they overlap.) Let \( I_n = U_n \cap V_n \), and let \( J_n = T_n - (U_n \cup V_n) \), that is, the set of elements of \( T_n \) that are not elements of \( U_n \cup V_n \).

It follows from our definitions that the following two propositions hold.

**Proposition 2.** The set \( I_n \subset T_n \) is the set of triples \((A, B, C)\) such that

- \( B \) is empty and \( C \) has both a point strictly above the horizontal axis and a point strictly below the horizontal axis.

**Proposition 3.** The set \( J_n \subset T_n \) is the set of triples \((A, B, C)\) such that

1. the path \( B \) ends with a down step and the path \( C \) is not empty, and the path \( C \) is entirely below the horizontal axis, or
2. the path \( B \) ends with an up step and the path \( C \) is not empty, and the path \( C \) is entirely above the horizontal axis.

Fortunately, there is a simple bijection mapping \( J_n \) to \( I_n \). Let \((A, B, C)\) ∈ \( J_n \), and let us assume without loss of generality that \((A, B, C)\) satisfies the first criterion of Proposition 3, that is, \( C \) is not empty, \( C \) is entirely below the horizontal axis, and \( B \) ends with a down step. Let \( z(A, B, C) = (A, \emptyset, BC) \), that is, we get \( z(A, B, C) \) by prepending \( C \) by \( B \), and thus making the middle path of the image empty. Note that \( z(A, B, C) \in I_n \), since the middle path of \( z(A, B, C) \) is empty, and its last path, \( BC \) has a point strictly above the horizontal axis (since \( B \) ends with a down step, and a point strictly below the horizontal axis, since \( C \) has such a point. If \((A, B, C)\) satisfies the second criterion of Proposition 3, the definition of \( z(A, B, C) \) is identical to the above, and the proof of the fact that \( z(A, B, C) \in I_n \) is analogous.

See Figure 6 for an illustration.

![Figure 6](image-url)
Lemma 4. The map $z : J_n \to I_n$ defined above is a bijection.

Proof. Let $(P, \emptyset, Q) \in I_n$. By definition, $R$ has a point strictly below, and a point strictly above, the horizontal axis. So there is at least one point that is common to $R$ and the horizontal axis at which $R$ crosses the horizontal axis, from one side to the other. Let $Y$ be the rightmost such point. Then the unique element $z^{-1}(P, \emptyset, Q) = (A, B, C)$ can be found by setting $A = P$, setting $C$ to be the part of $Q$ between $Y$ and the end of $Q$, and setting $B$ to be the part of $Q$ between the start of $Q$ and $Y$. □

Figure 7 illustrates the map $z^{-1}$.

3.4. The Composite Bijection. We are now in position to define the bijection that was our goal to construct in this section.

Definition 1. Let $n$ be any positive integer. Let $(A, B, C) \in T_n$, and let $g(A, B, C)$ be equal to

- $r(A, B, C)$ if $(A, B, C) \in R_n$,
- $s(A, B, C)$ if $(A, B, C) \in U_n$,
- $t(A, B, C)$ if $(A, B, C) \in V_n - U_n$,
- $t(z(A, B, C))$ if $(A, B, C) \in J_n$.

Theorem 2. The map $g : T_n \to D_n$ given in Definition 7 is a bijection.

Proof. We show that $g$ has an inverse. Let $(H, X) \in D_n$. If $(H, X) \in D_n^0$, then $g^{-1}(H, X) = r^{-1}(H, X)$. If $(H, X) \in D_n^+$, then $g^{-1}(H, X) = s^{-1}(H, X)$. Finally, if $(H, X) \in D_n^-$, then we need to consider $t^{-1}(H, X)$. If $t^{-1}(H, X) \in V_n - U_n$, then $g^{-1}(H, X) = t^{-1}(H, X)$, and if $t^{-1}(H, X) \in I_n = U_n \cap V_n$, then $g^{-1}(H, X) = z^{-1}(t^{-1}(H, X))$. This shows that each element of $D_n$ has exactly one preimage under $g$, proving that $g$ is a bijection. □
References

[1] D. Kleitman, A note on some subset identities. *Studies in Appl. Math.* 54 (1975), no. 4, 289–292.

[2] R. Stanley, Enumerative Combinatorics, Volume 1, second edition, Cambridge University Press, 1999.

[3] M. Sved, Counting and Recounting: The aftermath. *The Mathematical Intelligencer* 6 (1984), no. 4, 44–45.

Miklós Bóna, Department of Mathematics, University of Florida, 358 Little Hall, PO Box 118105, Gainesville, FL 32611–8105 (USA)