

A category for bijective combinatorics

Peter G. Doyle

Version 16 dated 20 June 2019
No Copyright*

Abstract

The category of matchings between finite sets extends to the category of cobordisms of signed sets. A chain of cobordisms that starts and ends with unsigned sets A and B yields a matching from A to B . This is a convenient way to package the involution principle of Garsia and Milne, which reveals itself to have little to do with involutions.

1 Introduction

As observed in passing by Conway and Doyle [2, p. 23], and doubtless by others before them, bijective combinatorics can be viewed as cobordism theory for oriented 0-dimensional manifolds. We develop this approach, with a view to clarifying the role of the involution principle of Garsia and Milne [5, 6].

There will be nothing new here, beyond notation. Cobordism theory dates from the 1950s, but its 0-dimensional manifestations can be seen in what is now the standard proof of the Cantor-Schroeder-Bernstein equivalence theorem, given by Koenig [7] in 1906. (See Appendix B.) And some will see the origins even further back in the mists of time. The application to combinatorics is implicit in Picciotto [8], and hardly different from the approach taken in texts like Stanton and White [9], and indeed the papers Garsia and Milne. In the end, it all comes down to subtraction.

*The authors hereby waive all copyright and related or neighboring rights to this work, and dedicate it to the public domain. This applies worldwide.

2 Notation

We deal with *matchings* (bijections) between finite sets. To emphasize this, we write

$$f : A \Rightarrow B$$

(‘ f matches A to B’), or

$$A \xRightarrow{f} B .$$

We write composition in natural order, using the symbol \triangleleft , pronounced ‘then’:

$$(f \triangleleft g)(x) = g(f(x)) .$$

We write $X + Y$ for the disjoint union

$$X + Y = X \times \{0\} \cup Y \times \{1\} ,$$

and adopt all the usual type coercions (‘abuses of notation’), so that

$$X \subset X + Y; X + Y = Y + X; (X + Y) + Z = X + (Y + Z) ,$$

etc. We can also take the disjoint union of matchings: If

$$f : A \Rightarrow B$$

and

$$g : C \Rightarrow D$$

then

$$f + g : A + B \Rightarrow C + D .$$

3 Simple subtraction

Theorem 1 (Simple subtraction) *If*

$$f : A + C \Rightarrow B + C$$

then

$$\text{cancel}(C, f) : A \Rightarrow B ,$$

where

$$\text{cancel}(C, f)(a) = \text{nestuntil}(\lambda x. x \notin C, f)(f(a))$$

and

$$\text{nestuntil}(\text{test}, f) = \lambda x. \text{if test}(x) \text{ then } x \text{ else } \text{nestuntil}(\text{test}, f)(f(x)) .$$

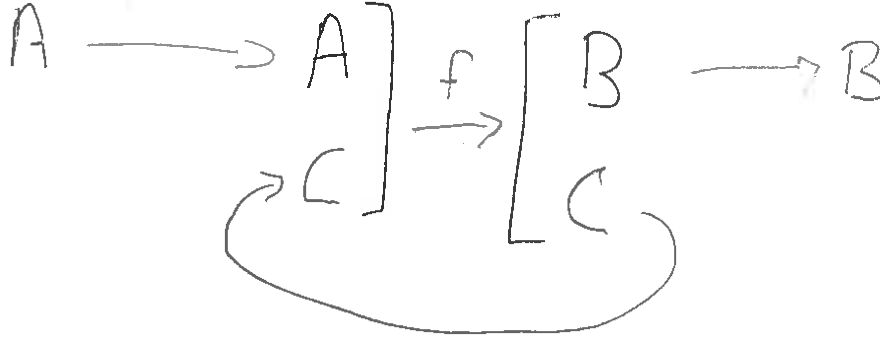


Figure 1: Simple subtraction.

Proof. Figure 1 shows the idea. The pidgin λ -calculus used to define $\text{cancel}(C, f)$ just means that, given a , we start with $f(a)$ and keep applying f until the result escapes from C . This happens eventually because C is finite. This escape mechanism applies to any function $f : A + C \rightarrow B + C$ to yield a function $\text{cancel}(C, f) : A \rightarrow B$. If f is an injection, so is $\text{cancel}(C, f)$; if f is a surjection, so is $\text{cancel}(C, f)$. ♠

Aside. We don't need A, B, C to be finite here: A and B may be infinite, and C need only be 'Dedekind-finite'—but let's not open that can of worms.

Corollary 2 (Repeated subtraction) *Let*

$$f : A + (C \cup D) \rightarrow B + (C \cup D)$$

where C, D are finite sets, not necessarily disjoint. Then

$$\text{cancel}(D, \text{cancel}(C, f)) = \text{cancel}(C \cup D, f) = \text{cancel}(C, \text{cancel}(D, f)) : A \Rightarrow B \quad .$$

Proof. If you persist in escaping from C until you have left D , to an outside observer your behaviour is indistinguishable from escaping from $C \cup D$. ♠

Here ends the math. The rest is bookkeeping.

4 The cobordism category

We've been working in the category whose morphisms are matchings of finite sets. Let's call this category **NSet**. Paralleling the extension of the natural numbers **N** to the integers **Z**, we are going to extend **NSet** to the category **ZSet** whose morphisms are cobordisms of signed sets.

4.1 Signed sets

Define a *signed set* to be an ordered pair $A = \langle A^+, A^- \rangle$ of finite sets. We identify unsigned sets as signed sets A for which $A^- = 0$, where 0 is the empty set, so that $\langle A^+, 0 \rangle = A^+$, and in particular $\langle 0, 0 \rangle = 0$. Soon we'll be writing $A = A^+ - A^-$, which is how we want to think of it.

Write

$$|A| = A^+ + A^- \quad ,$$

and define

$$A \subset B \iff |A| \subset |B| \iff A^+ \subset B^+ \wedge A^- \subset B^- \quad .$$

If $A \subset B$, define

$$B \setminus A = \langle B^+ \setminus A^+, B^- \setminus A^- \rangle \subset B \quad .$$

Define

$$A \leq B \iff A^+ \subset B^+ \wedge A^- \supset B^- \quad ,$$

and observe that being unsigned is the same as being (weakly) positive:

$$A \geq 0 \iff A \text{ is unsigned} \iff A = A^+ \iff A = |A| \quad .$$

(This way of designating unsigned sets is the only use we will make of this partial order.)

Define sum, negation, and difference of signed sets in the obvious ways:

$$A + B = \langle A^+ + B^+, A^- + B^- \rangle \quad ;$$

$$-A = \langle A^-, A^+ \rangle \quad ;$$

$$A - B = A + (-B) = \langle A^+ + B^-, A^- + B^+ \rangle \quad .$$

Now for a signed set A we can write

$$A = \langle A^+, A^- \rangle = A^+ - A^- \quad .$$

Everything works as expected, except that we can't replace $A - A$ with 0 , or vice versa.

4.2 Cobordisms

A *cobordism* is a matching of signed sets, represented by a triple $(|f|, A, B)$, where A, B are signed sets, and

$$|f| : A^+ + B^- \Rightarrow A^- + B^+ \quad .$$

We use the same notation for cobordisms as for matchings, writing

$$f : A \Rightarrow B$$

and saying ‘ f matches A to B ’ (or ‘ f is a cobordism from A to B ’, if we want to emphasize that we’re dealing with signed sets).

If

$$A \xRightarrow{f} B \xRightarrow{g} C \quad ,$$

we have matchings

$$\begin{aligned} |f| : A^+ + B^- &\Rightarrow A^- + B^+ \quad , \\ |g| : B^+ + C^- &\Rightarrow B^- + C^+ \quad . \end{aligned}$$

These combine to give a matching

$$|f| + |g| : A^+ + |B| + C^- \Rightarrow A^- + |B| + C^+ \quad .$$

We define

$$f \triangleleft g : A \Rightarrow C$$

by setting

$$|f \triangleleft g| = \text{cancel}(|B|, |f| + |g|) : A^+ + C^- \Rightarrow A^- + C^+ \quad .$$

(Cf. The ‘Bread Lemma’ of Picciotto [8, p. 25, Lemma 2]).

From Corollary 2 (repeated subtraction) we get

Corollary 3 (Associativity for composition of cobordisms) *Suppose*

$$A \xRightarrow{f} B \xRightarrow{g} C \xRightarrow{h} D.$$

Then

$$((f \triangleleft g) \triangleleft h) = (f \triangleleft (g \triangleleft h)) = f \triangleleft g \triangleleft h : A \Rightarrow D \quad ,$$

where

$$|f \triangleleft g \triangleleft h| = \text{cancel}(|B| + |C|, |f| + |g| + |h|) : A^+ + D^- \Rightarrow A^- + D^+ \quad . \quad \spadesuit$$

Corollary 4 (Chain associativity) *If*

$$A_0 \xRightarrow{f_{0,1}} A_1 \xRightarrow{f_{1,2}} \dots \xRightarrow{f_{n-2,n-1}} A_{n-1} \xRightarrow{f_{n-1,n}} A_n$$

then

$$|f_{0,1} \triangleleft \dots \triangleleft f_{n-1,n}| = \text{cancel}(|A_1| + \dots + |A_{n-1}|, |f_{0,1}| + \dots + |f_{n-1,n}|) \quad .$$

In particular, if $A_0, A_n \geq 0$ then

$$\text{cancel}(|A_1| + \dots + |A_{n-1}|, |f_{0,1}| + \dots + |f_{n-1,n}|) : A_0 \xRightarrow{\spadesuit} A_n \quad \spadesuit$$

(Cf. Picciotto [8, p. 26, Lemma 3].) We've written this out to emphasize that composing a chain of cobordisms requires only a single application of subtraction.

Of course we also have identity cobordisms

$$\text{id}(A) : A \xRightarrow{\spadesuit} A$$

with

$$|\text{id}(A)| := \text{id}(|A|) \quad ,$$

where in the second instance id denotes the identity in \mathbf{NSet} . So we have ourselves a category, which we call \mathbf{ZSet} .

Just as unsigned sets correspond to signed sets A with $A \geq 0$, matchings correspond to cobordisms

$$f : A \xRightarrow{\spadesuit} B$$

where $A, B \geq 0$, so that

$$|f| : A \xRightarrow{\spadesuit} B \quad .$$

This correspondence is natural ('functorial'): If $A, B, C \geq 0$ and

$$A \xRightarrow{f} B \xRightarrow{g} C$$

then

$$|f \triangleleft g| = |f| \triangleleft |g| : A \xRightarrow{\spadesuit} C \quad .$$

So we can identify a matching as a cobordism f for which $f = |f|$, just as an unsigned set is a signed set for which $A = |A|$.

4.3 Arithmetic with cobordisms

We can add, negate, and subtract cobordisms, bearing in mind that negating a cobordism reverses the direction of the arrow: If

$$f : A \Rightarrow B$$

then

$$-f : -B \Rightarrow -A \quad ,$$

with

$$|-f| = |f| : B^- + A^+ \Rightarrow B^+ + A^- \quad .$$

Now we can write the identity cobordism as

$$\text{id}(A) = \text{id}(A^+) - \text{id}(A^-) : A \Rightarrow A \quad .$$

Closely related to the identity are the creation and destruction cobordisms

$$\text{create}(A) : 0 \Rightarrow A - A$$

and

$$\text{destroy}(A) = -\text{create}(A) : A - A \Rightarrow 0 \quad ,$$

where

$$|\text{create}(A)| = |\text{destroy}(A)| = |\text{id}(A)| = \text{id}(|A|) \quad .$$

Observe that

$$\text{create}(A) = \text{create}(-A) = \text{create}(|A|) : 0 \Rightarrow A - A = (-A) - (-A) = |A| - |A| \quad .$$

More generally, for any

$$\phi : A \Rightarrow B$$

we define corresponding creation and destruction morphisms

$$\text{create}(\phi) : 0 \Rightarrow B - A$$

and

$$\text{destroy}(\phi) = -\text{create}(\phi) : A - B \Rightarrow 0$$

with

$$|\text{create}(\phi)| = |\text{destroy}(\phi)| = |\phi| : A^+ + B^- \Rightarrow A^- + B^+ \quad .$$

We have

$$\text{create}(\phi) = \text{create}(-\phi) = \text{create}(|\phi|) : 0 \Rightarrow B - A \quad .$$

5 Signed subtraction

With this machinery in place, we immediately get

Corollary 5 (Subtraction) *For any signed sets A, B, C, D , if*

$$f : A + C \Rightarrow B + D$$

and

$$g : D \Rightarrow C$$

then

$$(\text{id}(A) + \text{create}(C)) \triangleleft (f - g) \triangleleft (\text{id}(B) - \text{create}(D)) : A \Rightarrow B \quad .$$

Proof.

$$-g : -C \Rightarrow -D \quad ,$$

so

$$f - g : A + C - C \Rightarrow B + D - D \quad ,$$

so

$$A \xrightarrow{\text{id}(A) + \text{create}(C)} \Rightarrow A + C - C \xrightarrow{f - g} \Rightarrow B + D - D \xrightarrow{\text{id}(B) - \text{create}(D)} \Rightarrow B \quad . \spadesuit$$

Figure 2 shows a diagram. Taking $C = D$, $g = \text{id}(C)$ we recover Figure 1, so it might seem that we've made little progress beyond Theorem 1, and in a sense this is very true. Of course we now have subtraction working for signed sets, as emphasized in the exploded view of Figure 3. But mainly, we've just taken simple subtraction, dressed it up, and called it 'composition of morphisms'.

6 The involution principle

Corollary 6 (The involution principle) *For any signed sets X, Y and $A, B \subset X$, if*

$$\phi : Y \Rightarrow X \setminus A$$

and

$$\psi : X \setminus B \Rightarrow Y$$

then

$$(\text{id}(A) + \text{create}(\phi)) \triangleleft (\text{id}(B) - \text{create}(\psi)) : A \Rightarrow B \quad .$$

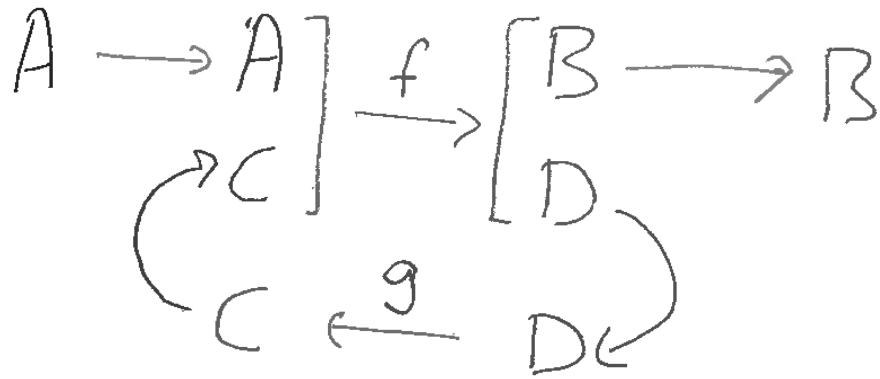


Figure 2: Subtraction.

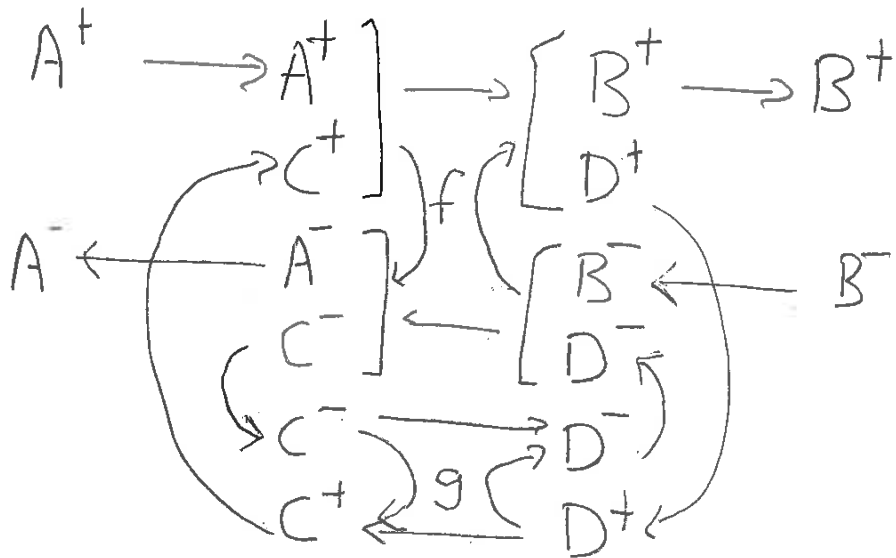


Figure 3: Subtraction (exploded view).

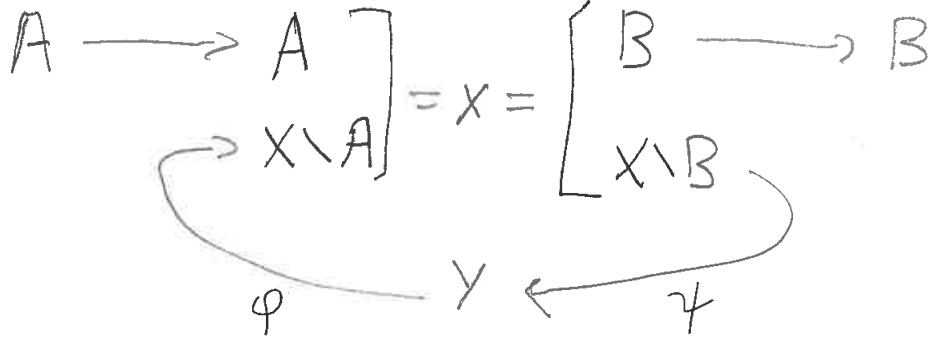


Figure 4: The involution principle.

Proof.

$$\text{create}(\phi) : 0 \Rightarrow (X \setminus A) - Y$$

and

$$-\text{create}(\psi) : (X \setminus B) - Y \Rightarrow 0$$

so

$$A \xrightarrow{\text{id}(A)+\text{create}(\phi)} A+(X \setminus A)-Y = X-Y = B+(X \setminus B)-Y \xrightarrow{\text{id}(B)-\text{create}(\psi)} B \quad \spadesuit$$

Figure 4 shows the diagram; Figure 5 shows the exploded view.

If we restrict to unsigned sets $X, Y \geq 0$ we recover the Garsia-Milne involution principle.

Corollary 7 (The Garsia-Milne involution principle) *If $A \subset X \geq 0$, $B \subset Y \geq 0$, $\phi : Y \Rightarrow X \setminus A$, $\psi : X \setminus B \Rightarrow Y$ then*

$$h = (\text{id}(A) + \text{create}(\phi)) \triangleleft (\text{id}(B) - \text{create}(\psi)) : A \Rightarrow B \quad ,$$

where

$$h(a) = \text{nestuntil}(\lambda x. x \in B, \lambda x. \phi(\psi(x)))(a) \quad . \quad \spadesuit$$

Formulated in this way, the involution principle has no need of involutions. If we want them nevertheless, we can manufacture them, as long as we're

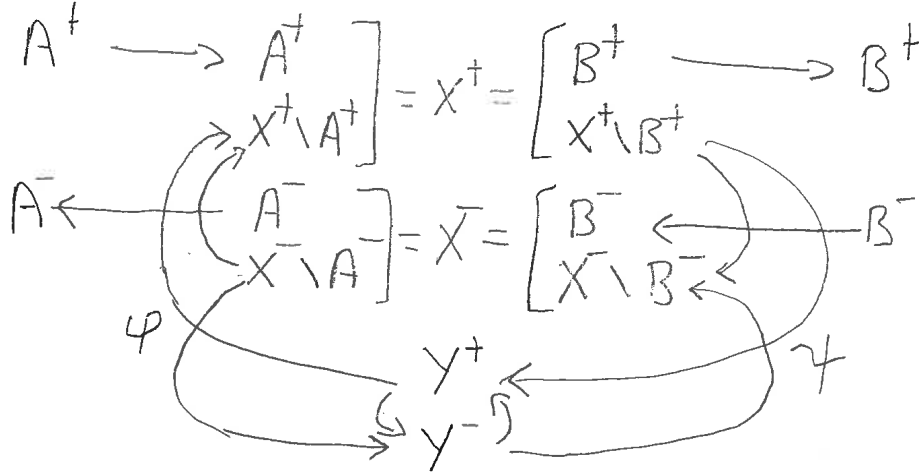


Figure 5: The involution principle (exploded view).

willing to take inverses of matchings. (Have you noticed that we've been studiously avoiding doing this?) From

$$\phi : Y \rightleftharpoons X \setminus A$$

we get

$$\phi^{-1} : X \setminus A \rightleftharpoons Y \quad ,$$

so

$$\Phi = \text{id}(A) + \phi + \phi^{-1} : X + Y = A + Y + (X \setminus A) \rightleftharpoons A + (X \setminus A) + Y = X + Y \quad ,$$

with

$$\Phi \triangleleft \Phi = \text{id}(X + Y) \quad .$$

Likewise,

$$\psi : X \setminus B \rightleftharpoons Y \quad ,$$

$$\psi^{-1} : Y \rightleftharpoons X \setminus B \quad ,$$

$$\Psi = \text{id}(B) + \psi + \psi^{-1} : X + Y = B + (X \setminus B) + Y \rightleftharpoons B + Y + (X \setminus B) = X + Y \quad ,$$

with

$$\Psi \triangleleft \Psi = \text{id}(X + Y) \quad .$$

Now we have

$$h(a) = \text{nestuntil}(\lambda x. \Psi(x) = x, \lambda x. \Phi(\Psi(x)))(a) \quad .$$

In practice, ϕ and ψ will often naturally arise as restrictions of involutions Φ and Ψ . And when this happens, exchanging Φ and Ψ gets us the inverse matching:

$$h^{-1}(b) = \text{nestuntil}(\lambda x. \Phi(x) = x, \lambda x. \Psi(\Phi(x)))(b) \quad .$$

Still, it will be helpful to recognize that fundamentally, the involution principle has little to do with involutions.

A Sidestepping division

The paper of Garsia and Milne [6] is one of the landmarks of bijective combinatorics. Beyond the specific application to the Rogers-Ramanujan identities, this work was a triumph for the null hypothesis that where there is algebra, there is combinatorics; it introduced combinatorialists to subtraction; and it showed the virtue of working with signed sets, which we've been touting.

Here we call attention to yet another aspect of their work, which was the way they avoided having to divide.

At its most basic, the problem of division is this. Suppose A, B, C are finite sets, with $C \neq 0$. From

$$f : A \times C \Rightarrow B \times C \quad ,$$

we want to produce

$$h : A \Rightarrow B \quad .$$

Rephrased for signed sets, from

$$f : A \times C \Rightarrow 0 \quad ,$$

we want to produce either

$$g : C \Rightarrow 0$$

or

$$h : A \Rightarrow 0 \quad .$$

Subtraction is straight-forward, but division is not. There are situations where division is needed, and techniques that will make it work. (Cf. Feldman and Propp [4]; Doyle and Qiu [3]; Bajpai and Doyle [1].) Garsia and Milne's insight was that, when working with generating functions, multiplying by the reciprocal may obviate the need to divide.

To give the idea, suppose we have generating function F, G that we wish to show are equal. We have a bijection showing $F \cdot H = G \cdot H$, and we want to derive a bijection showing $F = G$. The Garsia-Milne approach is to multiply by the reciprocal power series $K = H^{-1}$. A bijection showing $H \cdot K = 1$ yields bijections showing that $F = F \cdot H \cdot K$ and $G = G \cdot H \cdot K$. Now we have a chain of bijections showing

$$F = F \cdot H \cdot K = G \cdot H \cdot K = G \quad .$$

In each degree, we have a chain of bijections of signed sets, beginning and ending with unsigned sets. By subtraction, we get a matching between the terms of F and G .

This clever way to sidestep division was a key aspect of Garsia and Milne's work.

B Koenig's proof

Here is Koenig's proof of the Cantor-Schroeder-Bernstein equivalence theorem, reprinted from [7], with some trivial misprints corrected.

SÉANCE DU 9 JUILLET 1906.

ANALYSE MATHÉMATIQUE — *Sur la théorie des ensembles.*
Note de **M. Jules Koenig**, présentée par M. H. Poincaré.

La nouvelle démonstration du théorème d'équivalence de M. Cantor que je veux donner dans ces lignes a, comme je crois, une importance assez grande, vu la discussion actuelle sur les fondements de la logique, de l'arithmétique et de la théorie des ensembles. Je ne voulais la donner que dans l'exposition de la *Logique synthétique*, que j'espère publier bientôt et que j'ai déjà donnée dans mon cours de cette année. Mais l'intérêt qu'on prend aujourd'hui à ces choses me fait publier cette Note.

La critique spirituelle et profonde de M. Poincaré (voir la *Revue de Métaphysique et de Morale*, mai 1906) est irréfutable, à ce que je crois, dans ses parties négatives. Ce que nous possédons jusqu'à présent était peut-être nécessaire pour le développement de la nouvelle science logique; mais certainement cela ne donne pas ce que nous cherchons: les bases de cette nouvelle science.

Quant au théorème cité, énoncé pour la première fois par M. Cantor et démontré après par MM. Bernstein, Schroeder et Zermelo, il nous faudrait le mettre en évidence, sans employer le concept de nombre.

De plus nous devrions éviter le principe d'induction complète, pendant que, comme M. Poincaré l'a remarqué bien justement, toutes les démonstrations publiées jusqu'ici en font emploi. (Quant au concept de nombre, il est bien vrai que nous devons le construire nous-mêmes. Il y en a bien quelque chose dans l'intuition immédiate, un *fait vécu* ou une *expérience*; mais ce résidu est de toute nécessité.)

Le théorème d'équivalence est un théorème d'intuition. Pour démontrer cela j'emploierai la terminologie de M. Cantor; mais en soulignant en même temps qu'une exposition plus étendue et plus précise ne pourrait plus se servir des mots *ensemble*, etc.

Soient X, Y des ensembles déterminés, X_1, Y_1 des ensembles partiels de X et de Y respectivement. Nous devons démontrer que, étant $X \sim Y_1$ et

$Y \sim X_1$, nous aurons toujours $X \sim Y$.

La proposition $X \sim Y_1$ signifie la supposition de la loi (I) suivante:

Un élément quelconque x de X détermine un et un seul élément y de Y ; donc cet y détermine aussi le x correspondant. Mais il y a un ou plusieurs éléments de Y qui ne figurent pas dans cette loi.

De même la proposition $Y \sim X_1$ signifie la supposition d'une loi (II), qu'il serait superflu de détailler encore.

Prenons donc un élément quelconque x_1 de X ; après (I), il nous donne un élément déterminé y_1 de Y_1 ; cet élément y_1 nous donne, puis par la loi (II), un élément déterminé x_2 de X_1 , etc. En faisant cela, nous ne *comptons* pas; il n'y a là qu'un emploi des signes $1, 2, \dots$ pour distinguer les éléments de X . Mais les concepts *suivre* et *suite* doivent bien être acceptés comme concepts logiques définitifs.

Ainsi la suite

$$x_1 y_1 x_2 y_2 \dots$$

peut toujours être continuée à droite, mais pas toujours à gauche. Si x_1 est un élément de X_1 , la loi (II) donne un élément y_0 , qui précède immédiatement x_1 ; mais si x_1 est un élément de X , qui ne se trouve pas dans X_1 , la suite ne pourra plus être continuée à gauche.

On voit donc que les cas possibles sont les suivants:

La suite commence avec un élément de X . La suite commence avec un élément de Y . La suite peut toujours être prolongée à gauche.

Les éléments x'_1 et x''_1 de X nous donnent ainsi deux suites correspondantes:

$$x'_1 y'_1 x'_2 y'_2 \dots \tag{1}$$

$$x''_1 y''_1 x''_2 y''_2 \dots \tag{2}$$

S'il y a un élément commun dans les suites (1) et (2), l'élément qui le suit est déterminé par la loi (I), en conséquence il sera le même dans les suites (1) et (2), de même le précédent s'il y en a.

C'est-à-dire: Un élément quelconque de X détermine toujours la suite correspondante. Il n'est pas nécessaire de détailler le cas spécial d'une suite *périodique*. C'est évident, qu'une suite périodique peut toujours être prolongée à gauche.

La loi d'équivalence, dont l'expression est $X \sim Y$, se trouve déterminée de fait par ces considérations.

Soit \bar{x} un élément quelconque de X ; nous avons l'instruction pour la formation de la suite correspondante. Si cette suite commence avec un élément

de X , ou si elle peut être continuée à gauche, nous choisirons comme élément correspondant à \bar{x} dans Y l'élément qui le suit dans la suite. Mais, si la suite commence avec un élément de Y , nous prendrons comme élément correspondant dans Y celui qui précède \bar{x} immédiatement.

Ainsi l'équivalence $X \sim Y$ est fixée. L'intuition pure nous mène à reconnaître son *existence*.

Il va sans dire que cette exposition a encore beaucoup d'inconvénients; parce que nous n'avons pas discuté à fond les concepts logiques qui s'y trouvent. Telle est aussi l'expression à *droite* ou à *gauche*.

References

- [1] Prajeet Bajpai and Peter G. Doyle. Equivariant division, 2016, arXiv:1704.04089 [math.CO]. <http://arxiv.org/abs/1704.04089>.
- [2] Peter G. Doyle and John Horton Conway. Division by three, 1994, arXiv:math/0605779 [math.LO]. <http://arxiv.org/abs/math/0605779>.
- [3] Peter G. Doyle and Cecil Qiu. Division by four, 2015, arXiv:1504.01402 [math.LO]. <http://arxiv.org/abs/1504.01402>.
- [4] David Feldman and James Propp. Producing new bijections from old. *Adv. Math.*, 113:1–44, 1995.
- [5] A. M. Garsia and S. C. Milne. Method for constructing bijections for classical partition identities. *Proc. Natl. Acad. Sci. USA*, 78:2026–2028, 1981.
- [6] A. M. Garsia and S. C. Milne. A Rogers-Ramanujan bijection. *J. Combin. Theory, Series A*, 31:289–339, 1981.
- [7] Jules Koenig. Sur le théorie des ensembles. *C. R. Héb. Acad. Sci. Paris*, 143:110–112, 1906. Reprinted here as Appendix B.
- [8] Sally Picciotto. *How to encode a tree*. PhD thesis, UCSD, 1999, arXiv:1710.08463 [math.CO]. <http://arxiv.org/abs/1710.08463>.
- [9] Dennis Stanton and Dennis White. *Constructive Combinatorics*. Springer, 1986.