Why all rings should have a 1

Bjorn Poonen
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, MA 02139-4307, USA
poonen@math.mit.edu

June 15, 2018

Should the definition of ring require the existence of a multiplicative identity 1?

Emmy Noether, when giving the modern axiomatic definition of a commutative ring, in 1921, did not include such an axiom [15, p. 29]. For several decades, algebra books followed suit [16, §3.1], [18, I.§5]. But starting around 1960, many books by notable researchers began using the term "ring" to mean "ring with 1" [7, 0.(1.0.1)], [14, II.§1], [17, p. XIV], [1, p. 1]. Sometimes a change of heart occurred in a single person, or between editions of a single book, always towards requiring a 1: compare [11, p. 49] with [13, p. 86], or [2, p. 370] with [3, p. 346], or [4, I.§8.1] with [5, I.§8.1]. Reasons were not given; perhaps it was just becoming increasingly clear that the 1 was needed for many theorems to hold; some good reasons for requiring a 1 are explained in [6].

But is either convention more *natural*? The purpose of this article is to answer yes, and to give a reason: existence of a 1 is a part of what associativity should be.

1 Total associativity

The usual associative law is about reparenthesizing *triples*: $(ab)c = a(bc)$ for all a, b, c. Why are there not also ring axioms about reparenthesizing n-tuples for $n > 3$? It is because they would be redundant, implied by the law for triples. The whole point of associativity is that it lets us assign an unambiguous value to the product of any finite sequence of two or more terms.

By why settle for "two or more"? Cognoscenti do not require *sets* to have two or more elements. So why restrict attention to sequences with two or more terms? Most natural would be to require *every* finite sequence to have a product, even if the sequence is of length 1 or 0. This suggests the following:

Definition. A product on a set A is a rule that assigns to each finite sequence of elements of A an element of A, such that the product of a 1-term sequence is the term.

Next, let us explain why the usual associative law is insufficient to regulate such products. The usual associative law, although it involves three elements at a time, is a condition on a *binary* operation; that is, it constrains only the 2-fold products. But the new definition of product provides a value also to the 4-term sequence $abcd$, and so far there is no axiom to require this value to be the same as $((ab)c)d$ built up using the product on pairs repeatedly. We need a stronger associativity axiom to relate all the products of various lengths. This motivates the following:

Definition. A product is totally associative if each finite product of finite products equals the product of the concatenated sequence (for example, $(abc)d(ef)$ should equal the 6-term product $abcdef$).

Note that the finite products in this definition are not required to involve two or more terms; indeed, the definition would be more awkward if it spoke only of "finite products of two or more finite products of two or more terms each".

As argued at the beginning of this section, a product is more natural than a binary operation, insofar as it does not assign preferential status to the number 2. Similarly, total associativity, although less familiar than associativity, is more natural in that a law applicable to all tuples is more natural than a law applicable only to triples; after all, the number 3 is not magical either. Hence the ring axioms should be designed so that they give rise to a totally associative product. Now the key point is the following theorem, whose proof will be sketched at the end of this section:

Theorem. *A binary operation extends to a totally associative product if and only if it is associative and admits an identity element.*

What?! Where did that identity element come from? The definition of totally associative implies the equations

$$\begin{aligned}(abc)d &= abcd \\ (ab)c &= abc \\ (a)b &= ab \\ ()a &= a.\end{aligned}$$

The last equation, which holds for any a, shows that the empty product $()$ is a left identity. Similarly, $()$ is a right identity, so $()$ is an identity element.

Thus the natural extension of associativity demands that rings should contain an empty product, so it is natural to require rings to have a 1. But occasionally one does encounter structures that satisfy all the axioms of a ring except for the existence of a 1. What should they be called? Happily, there is an apt answer, suggested by Louis Rowen [13, p. 155]: rng! (Other suggestions include pseudo-ring [5, I.§8.1] and (associative) $\mathbb{Z}$-algebra [6, Appendix A].)

As our reasoning explains and as Rowen's terminology suggests, it is better to think of a rng as a ring with something missing than to think of a ring with 1 as having something extra.

Sketch of proof of the theorem. Given a binary operation that extends to a totally associative product, the argument above shows that it admits an identity element, and the usual associative law $(ab)c = a(bc)$ follows too since total associativity implies that both sides equal the 3-term product abc.

Conversely, given a binary operation $*$ that is associative and admits a 1, define

$$a_1 a_2 \cdots a_n := \begin{cases} 1, & \text{if } n = 0; \\ a_1, & \text{if } n = 1; \\ (a_1 a_2 \cdots a_{n-1}) * a_n, & \text{(inductively) if } n \geq 2. \end{cases}$$

This is a product extending $*$, and an involved but standard inductive argument shows that the usual associative law for $*$ can be used over and over to reparenthesize any finite product of finite products into the product of the concatenated sequence; for the details, see [5, I.§1.2, Théorème 1, and §2.1]. In other words, this extension of $*$ is a totally associative product. $\square$

2 Counterarguments

Here we mention some arguments for *not* requiring a 1, in order to rebut them.

- "Algebras should be rings, but Lie algebras usually do not have a 1."

  Lie algebras, which are objects used in more advanced mathematics to study the group of invertible $n \times n$ real matrices and its subgroups [8], are usually not associative either. We require a 1 only in the presence of associativity. It is accepted nearly universally that ring multiplication should be associative, so when the word "algebra" is used in a sense broad enough to include Lie algebras, it is understood that algebras have no reason to be rings.

- "An infinite direct sum of nonzero rings does not have a 1."

  By definition, if $A_1, A_2, \ldots$ are abelian groups, say, then the direct product $\prod_{i=1}^{\infty} A_i$ is the set of tuples $(a_1, a_2, \ldots)$ with $a_i \in A_i$ for each i, while the direct sum $\bigoplus_{i=1}^{\infty} A_i$ is the subgroup consisting of those tuples satisfying the additional condition that there are only finitely many i for which a_i is nonzero. Direct sums are typically defined for objects like vector spaces and abelian groups, for which the set of homomorphisms between two given objects is an abelian group, for which quotients exist, and so on. Rings fail to have these properties, whether or not a 1 is required: the quotient of a ring by a subring is no longer a ring (there is no natural way to multiply elements of $\mathbb{R}/\mathbb{Z}$, for instance). So it is strange even to speak of a direct sum of rings. Instead one should speak of the direct product, which does have a 1, namely the tuple with a 1 in every position.

- "If a 1 is required, then function spaces like the space $C_c(\mathbb{R})$ of continuous functions $f\colon \mathbb{R} \to \mathbb{R}$ that are 0 outside some unspecified bounded interval will be disqualified."

  This is perhaps the hardest to rebut, given the importance of function spaces. But many such spaces are ideals in a natural ring (e.g., $C_c(\mathbb{R})$ is an ideal in the ring $C(\mathbb{R})$ of *all* continuous functions $f\colon \mathbb{R} \to \mathbb{R}$), and fail to include the 1 only because of some condition imposed on their elements. So one can say that they, like the direct sums above and like the rng of even integers, *deserve* to be ousted from the fellowship of the ring. In any case, however, these function spaces still qualify as $\mathbb{R}$-algebras.

Further counterarguments can be found in the preface to [9].

Rngs should not be banished completely, because there are applications for which rngs are more convenient than rings: see [12, pp. 31–36] and [10], for instance. The latter gave the first example of an infinite finitely generated group all of whose elements have finite order.

3 Implications

Once the role of the empty product is acknowledged, other definitions that seemed arbitrary become natural.

- A ring homomorphism $A \to B$ should respect finite products, so in particular it should map the empty product 1_A to the empty product 1_B.

- A subring should be closed under finite products, so it should contain the empty product 1.

- An ideal is prime if and only if its complement is closed under finite products. This explains why the unit ideal (1) in a ring is never considered to be prime: if (1) were a prime ideal, then its complement $\emptyset$ would be closed under finite products and in particular would contain the empty product 1; but $\emptyset$ does not contain 1.

- The argument that rings should have a 1 involved only one binary operation, multiplication, so the same argument explains also why monoids are more natural than semigroups. (A semigroup is a set with an associative binary operation, and a monoid is a semigroup with a 1.)

There are also applications of these ideas in higher mathematics. For example, there is the notion of a category, which has objects and morphisms satisfying certain axioms modeled on the properties of groups and homomorphisms. One of these axioms lets one compose morphisms $f\colon A \to B$ and $g\colon B \to C$ to produce $g \circ f$ when the target object of f (here called B) matches the source object of g. Another axiom asserts the existence of identity morphisms. Now we can see why the axiom about identity morphisms is natural: it arises as a special case of composing a chain of morphisms. More specifically, given objects $A_0, \ldots, A_n$ and a chain of morphisms

$$A_0 \xrightarrow{f_1} A_1 \xrightarrow{f_2} \cdots \xrightarrow{f_n} A_n,$$

one wants to be able to form the composition, even when $n = 0$, and in the $n = 0$ case the composition is the identity morphism from A_0 to A_0.

4 Final comments

It would be ridiculous to introduce the definition of ring to beginners in terms of totally associative products. But it is nice to understand why certain definitions should be favored over others.

Acknowledgements

I thank Keith Conrad, Greg Marks, Gigel Militaru, Johan Öinert, and Lance Small for many helpful comments. I thank also the referees for many suggestions that improved this note. This research was supported in part by National Science Foundation grants DMS-1069236 and DMS-1601946 and grants from the Simons Foundation (#340694 and #402472 to Bjorn Poonen). The Version of Record of this manuscript has been published and is available in *Mathematics Magazine* **92**:1, 55–62, `http://www.tandfonline.com/10.1080/0025570X.2018.1538714`.

References

[1] M. F. Atiyah and I. G. Macdonald, *Introduction to commutative algebra*, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1969. MR0242802 (39 #4129)

[2] Garrett Birkhoff and Saunders Mac Lane, *A survey of modern algebra*, Macmillan Co., New York, N. Y., 1953. Rev. ed. MR0054551 (14,939a)

[3] Garrett Birkhoff and Saunders Mac Lane, *A survey of modern algebra*, Third edition, The Macmillan Co., New York, 1965. MR0177992 (31 #2250)

[4] N. Bourbaki, *Éléments de mathématique. Part I. Les structures fondamentales de l'analyse. Livre II. Algèbre. Chapitre I. Structures algébriques*, Actual. Sci. Ind., no. 1144, Hermann, Paris, 1958 (French).

[5] N. Bourbaki, *Éléments de mathématique. Algèbre. Chapitres 1 à 3*, Hermann, Paris, 1970; English transl., *Algebra I. Chapters 1–3*, Elements of Mathematics (Berlin), Springer-Verlag, Berlin, 1998. Reprint of the 1989 English translation. MR1727844. MR0274237 (43 #2)

[6] K. Conrad, *Standard definitions for rings*, April 25, 2013. Available at `http://www.math.uconn.edu/~kconrad/blurbs/ringtheory/ringdefs.pdf`.

[7] A. Grothendieck, *Éléments de géométrie algébrique. I. Le langage des schémas*, Inst. Hautes Études Sci. Publ. Math. **4** (1960), 228. MR0217083 (36 #177a)

[8] Karin Erdmann and Mark J. Wildon, *Introduction to Lie algebras*, Springer Undergraduate Mathematics Series, Springer-Verlag London, Ltd., London, 2006. MR2218355

[9] B. J. Gardner and R. Wiegandt, *Radical theory of rings*, Monographs and Textbooks in Pure and Applied Mathematics, vol. 261, Marcel Dekker, Inc., New York, 2004. MR2015465

[10] E. S. Golod, *On nil-algebras and finitely approximable p-groups*, Izv. Akad. Nauk SSSR Ser. Mat. **28** (1964), 273–276 (Russian). MR0161878

[11] Nathan Jacobson, *Lectures in abstract algebra. Vol. I. Basic concepts*, D. Van Nostrand Co., Inc., Toronto, New York, London, 1951. MR0041102 (12,794c)

[12] Nathan Jacobson, *Lie algebras*, Dover Publications, Inc., New York, 1979. Republication of the 1962 original. MR559927

[13] Nathan Jacobson, *Basic algebra. I*, 2nd ed., W. H. Freeman and Company, New York, 1985. MR780184 (86d:00001)

[14] Serge Lang, *Algebra*, Addison-Wesley Publishing Co., Inc., Reading, Mass., 1965. MR0197234 (33 #5416)

[15] Emmy Noether, *Idealtheorie in Ringbereichen*, Math. Ann. **83** (1921), no. 1-2, 24–66, DOI 10.1007/BF01464225. MR1511996

[16] B. L. van der Waerden, *Algebra. Teil I*, Siebte Auflage. Heidelberger Taschenbücher, Band 12, Springer-Verlag, Berlin, 1966; English transl., *Algebra. Vol 1*, Translated by Fred Blum and John R. Schulenberger, Frederick Ungar Publishing Co., New York, 1970. MR0263582 (41 #8187a). MR0263581 (41 #8186)

[17] André Weil, *Basic number theory*, Die Grundlehren der mathematischen Wissenschaften, Band 144, Springer-Verlag New York, Inc., New York, 1967. MR0234930 (38 #3244)

[18] Oscar Zariski and Pierre Samuel, *Commutative algebra. Vol. 1*, Springer-Verlag, New York, 1975. With the cooperation of I. S. Cohen; Corrected reprinting of the 1958 edition; Graduate Texts in Mathematics, No. 28. MR0384768 (52 #5641)

5 Summary

We argue that the definition of ring should require the existence of a multiplicative identity 1 because this requirement is part of what associativity should be. We also address counterarguments and explore some implications of our argument.

6 Author information

Bjorn Poonen (MR Author ID 250625) is the Claude Shannon Professor of Mathematics at MIT. His research focuses mainly on number theory and algebraic geometry, and his expository writing earned him the MAA Chauvenet Prize. Nineteen mathematicians have completed a Ph.D. thesis under his guidance. He serves the mathematical community in various ways, for instance as managing editor of *Algebra & Number Theory*. Finally, he has been singing choral music for over 30 years.