Zorn's Bowels
The Mechanics of Choice and Recursion

John Peloquin

February 2008

Introduction

The Axiom of Choice (AC) in Zermelo-Fraenkel set theory (ZFC) allows one, roughly, to make infinitely many arbitrary choices. While it is a very powerful axiom, there are a number of reformulations[1] of it which can be more convenient to use directly in certain situations. Zorn's Lemma[2] is a popular such reformulation, restating AC in terms of the existence of maximal elements within (certain) partial orderings. It is useful because it is often straightforward to apply and allows one to avoid some of the messier mechanics of choice.

Despite its convenience as a tool, however, Zorn's Lemma can sometimes make a proof more opaque than it would be if it used choice directly. In particular, it can sometimes hide underlying similarity between a proof for an infinite case and the corresponding proof for a simpler (usually more intuitive) finite case. In this note, I show by example how this can occur, and how using choice (and recursion) directly can make a proof more transparent. I do not, of course, suggest abandoning Zorn's Lemma, but rather gaining insight into its use by way of the direct approach.[3]

A Problem from Algebra

In order to illustrate my point, I present an elementary problem from the theory of rings. For the reader unfamiliar with algebra, I first state some informal definitions.

A *ring* is simply a set of elements together with two binary operations defined on it, called *addition* and *multiplication*. These operations must satisfy a number of properties, the details of which I will not cover here. As an example, the integers form a ring under standard addition and multiplication; also, if m, n are positive integers, the set of all real $m \times n$ matrices forms a ring under matrix addition and multiplication. An *ideal* is just a subset of a ring that is closed under addition, and closed under multiplication by elements in the ring. For example, the subset of even

[1] By a *reformulation* of AC, I just mean a statement P in the language of set theory such that $\mathrm{AC} \leftrightarrow P$ is a theorem of ZFC − AC.

[2] The title of 'lemma' is for purely historical reasons.

[3] I am indebted to Leo Harrington for emphasizing so strongly the underlying simplicity of Zorn's Lemma in his undergraduate set theory course. See [6] for detailed notes.

integers forms an ideal in the integers. A *field* is a very special type of ring in which one can divide (by nonzero elements). Note that the integers do not form a field, but the rational numbers (fractions) under standard addition and multiplication do.

In the theory of (commutative) rings, ideals play a prominent role. A *maximal ideal* in a ring R is an ideal $I \neq R$ with no ideals other than R properly containing it. That is, if $I \subseteq R$ is an ideal, then I is maximal iff $I \neq R$ and for all ideals J with $I \subseteq J \subseteq R$, either $J = I$ or $J = R$. Maximal ideals are significant because the quotient R/I of a ring R with a maximal ideal I forms a field.

A natural question arises: do all (nontrivial) rings possess maximal ideals? To answer this question, we might first simplify things by considering the finite and infinite cases separately. In the finite case, a natural algorithm comes to mind for constructing a maximal ideal: simply start small and build up to one in a step by step manner. In more detail:

> Let R be the ring and start with the trivial ideal $I_0 = \{0\}$. We know $I_0 \neq R$ since R is nontrivial. Now consider the elements of R not in I_0. If there are no elements that can be adjoined to I_0 without generating the entire ring R,[4] then halt. On the other hand, if there is such an element r, then let I_1 be the ideal generated by adjoining r to I_0.
>
> Now repeat this process for $I_1, I_2, \ldots$

Since R is finite, we must eventually reach a step j at which we halt. At this point I_j cannot be extended without generating the entire ring R. Since at each step i we ensured $I_i \neq R$, we have $I_j \neq R$. Hence I_j satisfies both conditions for maximality, and is a maximal ideal. In particular, we proved R has a maximal ideal.

This proof is very intuitive, and the basic idea it embodies can be used to prove similar claims for finite groups, finite dimensional vector spaces, and so on. But it does not carry over immediately into the infinite case. More specifically, the above algorithm does not work in general for an infinite ring, since in such a ring it may be possible to construct ideals in the above manner and never reach a maximal ideal. After the 'first' infinitely many steps of the algorithm are performed, additional steps may still be required!

This is, of course, where Zorn's Lemma is typically introduced:

Zorn's Lemma. *Let $(P, \leq)$ be a partial ordering such that every chain in P has an upper bound. Then P has a maximal element.*

Recall that a *partial ordering* $\leq$ on a set P is a relation on P satisfying the following properties for all $x, y, z \in P$:

$$x \leq x$$
$$(x \leq y \wedge y \leq x) \rightarrow x = y$$
$$(x \leq y \wedge y \leq z) \rightarrow x \leq z$$

A *chain* in P is simply a subset $C \subseteq P$ such that for all $x, y \in C$, either $x \leq y$ or $y \leq x$ (in other words, a chain is *linearly ordered*). An *upper bound* of $C \subseteq P$ is an element

[4] By *generating* I mean with the smallest ideal containing all of I_0 and the adjoined element.

$y \in P$ such that $x \leq y$ for all $x \in C$. Finally, a *maximal element* in P is an element $x \in P$ such that there does not exist any $y \in P$ with $x < y$ (that is, $x \leq y$ and $x \neq y$).

To prove that an infinite ring R has a maximal ideal, we let P be the set of all proper ideals in R, partially ordered under set inclusion $\subseteq$. Then P is nonempty. If $C \subseteq P$ is any chain of proper ideals in R, it can be verified that the union $\bigcup C$ over C is also a proper ideal in R, and is an upper bound for C.[5] By Zorn's Lemma, P contains a maximal element, namely, a maximal ideal.

Thus we have answered our original question in the affirmative—all (nontrivial) rings, both finite and infinite, possess maximal ideals.

The Finite Case

Although we were able to answer our question, the proof for the infinite case leaves something to be desired. At first glance (or even second or third), it appears to be wildly different than the proof for the finite case, and it loses the intuitive appeal of the finitary algorithm. In addition, it seems a little bit like a magic trick, providing us with a maximal ideal but not making it obvious *why* one exists.

Can we obtain a more intuitive proof? In order to do so, we must first make our proof for the finite case more rigorous. In particular, we must make explicit the underlying elements involved in our finitary algorithm.

Examining our algorithm closely, we see that there are two basic processes at work: the choosing of elements which satisfy a given property (if they exist), and at each step the use of something defined at the previous step. The former is just an instance of (finite) choice;[6] the latter, an instance of recursion. We can make both of these factors explicit by translating our algorithm into a more rigorous (set theoretic) argument.

For what follows, we require some additional notation: if A and B are sets, then $A \subset B$ means $A \subseteq B$ and $A \neq B$. Also $A - B = \{x \in A \mid x \notin B\}$. For R a ring, I an ideal in R, and $r \in R$, $(I \cup \{r\})$ denotes the ideal generated by $I \cup \{r\}$—that is, the smallest ideal in R containing everything in I as well as the element r. Note that $I \subseteq (I \cup \{r\}) \subseteq R$. Also, ω denotes the set $\{0, 1, 2, \ldots\}$ of nonnegative integers.

> Suppose R is a nontrivial finite ring. Let $\mathscr{I}$ be the set of all ideals in R. Define a function F on a subset of $\mathscr{I}$ as follows: for each $I \in \mathscr{I}$, if there exists $r \in R - I$ such that $I \subset (I \cup \{r\}) \subset R$, choose such an r and define $F(I) = r$; otherwise, let $F(I)$ be undefined.
>
> Now recursively define a function $H : \omega \to \mathscr{I}$ as follows:
>
> $$H(0) = \{0\}$$
> $$H(n+1) = \begin{cases} \big(H(n) \cup \{F(H(n))\}\big) & \text{if } F(H(n)) \text{ is defined} \\ H(n) & \text{otherwise} \end{cases} \qquad (n \in \omega)$$

[5] Technically this is false if C is the empty chain, but I am being imprecise for clarity. Any element of P serves as an upper bound for the empty chain.

[6] This does not require the use of AC; more about that later.

For convenience, write $H_n = H(n)$.

It is verified by induction that H_n is a proper ideal in R for every $n \in \omega$. Similarly, $H_m \subseteq H_n$ for all $m \leq n$. Hence we have obtained an ascending chain of proper ideals in R.

We claim that there exist naturals $m < n$ such that $H_m = H_n$. Indeed, if this is not so, then H is a one-to-one mapping from ω into $\mathscr{I}$. But this is impossible since ω is infinite and $\mathscr{I}$ is finite (since R is finite).[7]

Let m be the least natural such that $H_m = H_n$ for some $n > m$. Note that $m < m+1 \leq n$, hence

$$H_m \subseteq H_{m+1} \subseteq H_n = H_m$$

so in particular $H_m = H_{m+1}$—that is, our chain stops growing at H_m.

We claim that H_m is maximal. Indeed, if this is not so, then there exists an ideal H such that $H_m \subset H \subset R$. Choose $r \in H - H_m$. Then

$$H_m \subset (H_m \cup \{r\}) \subseteq H \subset R$$

But then $F(H_m)$ must be defined, so by the definition of H, $H_m \subset H_{m+1}$—contradicting that $H_m = H_{m+1}$. It follows that H_m is indeed maximal. In particular, R has a maximal ideal.

This argument is certainly more technical than our original one, but it embodies the same basic underlying idea. In both arguments we start with the trivial ideal and build up to a maximal ideal in a step by step manner. The differences here are that elements are first chosen 'all at once' with a choice function F, and the original step by step algorithm is replaced with a recursive function H.

This more rigorous proof is valuable because it isolates the key components that will need to be replicated in the infinite case. First, we will need to be able to obtain a choice function for a possibly infinite set of ideals. Second, we will need a way to continue the recursion process 'beyond' ω. In the above proof, we see that the only reason the ascending chain of ideals collapses is that ω is larger than any finite set. But as far as infinite sets go, ω is tiny (in fact, it is smallest). We will need to be able to 'run through' the elements of any arbitrary infinite ring, and for this we require a more powerful type of recursion.

The General Case

To be able to extend recursion beyond ω, we must be able to count beyond ω. More specifically, we must have a number system which extends the nonnegative integers, allowing for the 'enumeration' of larger infinite sets.

In ZFC, the class Ω of *ordinals* extends ω in precisely this manner:

$$0,\ 1,\ 2,\ \ldots,\ \omega,\ \omega+1,\ \omega+2,\ \omega+3,\ \ldots$$

[7] This result is an instance of (a form of) the *Pigeonhole Principle*, which is a theorem of ZFC.

There are a number of different ways to define the ordinals, but the details of these definitions do not concern us. For our purposes, we can envision the ordinals as being 'constructed' in such a way that each ordinal is just the set of all previously constructed ordinals:[8]

$$\begin{aligned}
0 &= \emptyset \\
1 &= \{0\} = \{\emptyset\} \\
2 &= \{0,1\} = \{\emptyset, \{\emptyset\}\} \\
&\vdots \\
\omega &= \{0,1,2,\ldots\} \\
\omega+1 &= \{0,1,2,\ldots,\omega\} \\
\omega+2 &= \{0,1,2,\ldots,\omega,\omega+1\} \\
&\vdots
\end{aligned}$$

By definition there are three distinct types of ordinals: the *zero* ordinal, the *successor ordinals*, and the rest, called *limit ordinals*. A successor ordinal is any ordinal of the form $\alpha+1$ for some ordinal α. For any limit ordinal λ and any $\alpha < \lambda$, we must have $\alpha+1 < \lambda$. Note that all the positive integers are successor ordinals, as well as $\omega+1$, $\omega+2 = (\omega+1)+1$, etc. The first limit ordinal is ω, and others include

$$\omega+\omega,\ \omega+\omega+\omega,\ \ldots,\ \omega\cdot\omega,\ \omega\cdot\omega+\omega,\ \ldots,\ \omega\cdot\omega+\omega\cdot\omega,\ \ldots,\ \omega\cdot\omega\cdot\omega,\ \ldots,\ \omega^{\omega},\ \ldots$$

A particularly neat example is $\omega^{\omega^{\omega^{\cdot^{\cdot^{\cdot}}}}}$, where the exponentiation is ω-many times.

The class Ω is similar to the set ω in that there are always least elements:

Well Ordering on Ω. *Let Φ be any property defined on Ω and suppose there exists $\alpha \in \Omega$ such that $\Phi(\alpha)$ holds. Then there exists a least $\lambda \in \Omega$ such that $\Phi(\lambda)$ holds.*

From this we obtain:

Induction on Ω. *Let Φ be any property defined on Ω, and suppose the following is true for all $\alpha \in \Omega$:*

If $\Phi(\beta)$ holds for all $\beta < \alpha$, then $\Phi(\alpha)$ holds.

Then $\Phi(\alpha)$ holds for all $\alpha \in \Omega$.

It is intuitively evident from our informal construction that the ordinals 'go on forever'. In fact it can be formally proved in ZFC that the class Ω is not a set.[9] This is significant because it presents us with another similarity between Ω and ω: just as ω is larger than any *finite* set, Ω is 'larger' than any *arbitrary* set. This suggests that if we can perform recursions on Ω, we will be able to 'run through' the elements of any arbitrary infinite set.

It is a nontrivial fact that ZFC facilitates recursion on Ω:

[8] Note that this statement could not serve as a formal definition of an ordinal, for it would be deeply circular as it stands. A convenient definition, which is consistent with our informal construction, states that an ordinal is a well ordered set $(X,<)$ such that for all $y \in X$, $y = \{x \in X \mid x < y\}$.

[9] This result is known historically as the *Burali-Forti Paradox*, though it is not a paradox in ZFC.

Recursion on Ω. *Let $\mathscr{G}$ be an operation*[10] *defined on the class of all functions. Then there exists a (unique) operation $\mathscr{F}$ on Ω such that*

$$\mathscr{F}(\alpha) = \mathscr{G}(\mathscr{F}|_\alpha)$$

for all $\alpha \in \Omega$.

(Note that $\mathscr{F}|_\alpha$ refers to the restriction of $\mathscr{F}$ to α, that is, the unique function f on α such that $f(\beta) = \mathscr{F}(\beta)$ for all $\beta \in \alpha$.[11]) This result tells us that we can define recursive operations on Ω in precisely the way we defined a recursive function on ω above. By specifying how each value should be obtained from previous values, we acquire an operation that is defined on all ordinals.

In order to complete our proof for the infinite case, we still need a way to choose an arbitrary number of elements. This is provided by the Axiom of Choice in ZFC:

Axiom of Choice (AC). *Let C be a set of nonempty sets. Then there exists a function F on C such that $F(x) \in x$ for all $x \in C$.*

Note AC is not required when C is finite, since the existence of choice functions for finite collections can be proved from the other axioms.[12] But for infinite collections, in general, the existence of choice functions cannot be proved without AC, making it a fundamental tool in ZFC.

With choice and recursion in hand, we can finally construct our general proof:

> Suppose R is a nontrivial ring, and let $\mathscr{I}$ be the set of ideals in R. Define a function E on $\mathscr{I}$ by:
>
> $$E(I) = \{ r \in R \mid I \subset (I \cup \{r\}) \subset R \}$$
>
> for $I \in \mathscr{I}$. Now set $P = \{ I \in \mathscr{I} \mid E(I) \neq \emptyset \}$ and $C = \{ E(I) \mid I \in P \}$. By AC above, there exists a choice function F on C.
>
> Now define a recursive operation $\mathscr{F}$ on Ω as follows:
>
> $$\mathscr{F}(\alpha) = \begin{cases} \{0\} & \text{if } \alpha = 0 \\ \big(\mathscr{F}(\beta) \cup \{F(E(\mathscr{F}(\beta)))\}\big) & \text{if } \alpha = \beta + 1 \text{ and } \mathscr{F}(\beta) \in P \\ \mathscr{F}(\beta) & \text{if } \alpha = \beta + 1 \text{ and } \mathscr{F}(\beta) \notin P \\ \bigcup_{\beta<\alpha} \mathscr{F}(\beta) & \text{if } \alpha \text{ is a limit ordinal} \end{cases}$$
>
> For convenience, write $H_\alpha = \mathscr{F}(\alpha)$.
>
> It is verified by induction that H_α is a proper ideal in R for all $\alpha \in \Omega$. Similarly, $H_\alpha \subseteq H_\beta$ for all $\alpha \leq \beta$.

[10] An operation is like a function, except its domain is too large to form a set. Formally, an operation is a formula $\varphi(x, y)$ in the language of set theory such that for all sets a, there exists exactly one set b such that $\varphi(a, b)$ holds.

[11] By the Axiom of Replacement (among others) in ZFC, $\mathscr{F}|_\alpha$ is indeed a function.

[12] The proof proceeds by induction on finite cardinalities.

> Since there cannot exist a one-to-one operation from the class Ω into the set $\mathscr{I}$, there exists some α with $H_\alpha = H_\beta$ for some $\beta > \alpha$. As before, $\alpha < \alpha + 1 \leq \beta$, hence $H_\alpha = H_{\alpha+1}$.
>
> By definition of our operation then, we must have $H_\alpha \notin P$—that is, there does not exist $r \in R$ such that $H_\alpha \subset (H_\alpha \cup \{r\}) \subset R$. But then, as before, there cannot exist any ideal H with $H_\alpha \subset H \subset R$. Thus the only ideals containing H_α are H_α and R, which means H_α is maximal, and R has a maximal ideal.

While this proof contains a fair amount of technical machinery, it still embodies the basic idea from our original algorithm: given a ring, we start with the trivial ideal and build up to a maximal ideal in a step by step manner. The only difference is that we now have two different types of steps to carry out. At each successor step, we adjoin a new element to the previous ideal if this will not generate the entire ring. At each limit step, we simply take the union over all previously constructed ideals. The ascending chain produced must eventually stop growing since we are guaranteed to 'run out' of new elements to adjoin. At this point we arrive at a maximal ideal, establishing the existence claim.

I think that, despite its technical machinery, this proof is more transparent than the Zorn's Lemma proof. It naturally extends our original algorithm and illustrates *why* a maximal ideal exists. We see that in *any* ring it is possible to construct a chain of larger and larger proper ideals, and that such a chain must always eventually top out at a maximal element.

This proof also does something aesthetically which the Zorn's Lemma proof does not: it showcases the power and elegance of recursion on the ordinals. The ordinals were first developed by Cantor in large part to facilitate transfinite recursion. The above proof illustrates how useful, and beautiful, this concept really is.

Zorn's Lemma Revisited

Of course, I am not here suggesting the abandonment of Zorn's Lemma—as a tool, it is far too convenient to discard (not to mention deeply integrated into modern mathematics). The above problem is just one of countless examples in mathematics where maximal elements are involved, and it would be inefficient and tedious to continually replicate the machinery of choice and recursion for each such problem separately. But I think it is important to view Zorn's Lemma correctly (as a tool) and to understand conceptually how it works.

At this point it is fairly straightforward to see how to prove Zorn's Lemma using choice and recursion. If we are given a partial ordering $(P, \leq)$ such that every chain in P has an upper bound, then we can recursively construct an ascending chain of elements in P which must eventually top out at a maximal element.

In more detail:

> Let $(P, \leq)$ be a partial ordering such that every chain in P has an upper bound. We claim P has a maximal element.

Indeed, let $\mathscr{C}$ be the set of all chains in P. By AC, there exists a function $F : \mathscr{C} \to P$ such that $F(C)$ is an upper bound of C for all $C \in \mathscr{C}$. Similarly, there exists a function $G : P \to P$ such that $G(p) > p$ if p is not maximal, and $G(p) = p$ otherwise.

Fix $e \notin P$ and recursively define an operation $\mathscr{F}$ on Ω as follows:

$$\mathscr{F}(\alpha) = \begin{cases} G(F(\mathscr{F}[\alpha])) & \text{if } \mathscr{F}[\alpha] \in \mathscr{C} \\ e & \text{otherwise} \end{cases}$$

(Note $\mathscr{F}[\alpha] = \{\mathscr{F}(\beta) \mid \beta < \alpha\}$.) It is verified by induction that $\mathscr{F}[\alpha] \in \mathscr{C}$ for all $\alpha \in \Omega$, and hence $\mathscr{F}(\alpha) \in P$ for all $\alpha \in \Omega$. For convenience, write $p_\alpha = \mathscr{F}(\alpha)$.

Now $p_\alpha \leq p_\beta$ for all $\alpha \leq \beta$, and there must exist α with $p_\alpha = p_\beta$ for some $\beta > \alpha$. Since $\alpha < \alpha + 1 \leq \beta$, we have $p_\alpha = p_{\alpha+1}$.

Let $p = F(\{p_\beta \mid \beta \leq \alpha\})$. Then by definition of our recursion, we have

$$p_\alpha \leq p \leq G(p) = p_{\alpha+1}$$

Since $p_\alpha = p_{\alpha+1}$, we must have $G(p) = p$. By definition of G, this means p must be maximal in P.

Thus P has an maximal element, establishing our claim.

From this proof we see that Zorn's Lemma simply hides the machinery of choice and recursion. It allows us to obtain a maximal element by verifying a simple closure property because this property ensures that the above recursion can be done.

In the end, it is really a matter of personal preference whether to use the direct approach or Zorn's Lemma, or other equivalents, for these types of proofs. But in all cases it is useful to understand the simple underlying conceptual framework, for this can make even very technical proofs transparent.

References

[1] Bergman, George M. 'The Axiom of Choice, Zorn's Lemma, and all that' (course handout). UC Berkeley, Spring 1997. `http://math.berkeley.edu/~gbergman/grad.hndts/AC+Zorn+.ps`

[2] Devlin, Keith. *The Joy of Sets: Fundamentals of Contemporary Set Theory.* Springer, 1994.

[3] Dummit, David S. and Richard M. Foote. *Abstract Algebra, 3rd ed.* Wiley, 2003.

[4] Enderton, Herbert B. *Elements of Set Theory.* Academic Press, 1977.

[5] Hajnal, András and Peter Hamburger. *Set Theory.* Cambridge, 1999.

[6] Harrington, Leo. *Introduction to Set Theory* (course notes by John Peloquin). UC Berkeley, Spring 2007. `http://blargon.net/math/settheory.pdf`

[7] Hungerford, Thomas W. *Algebra.* Springer, 1974.