The Axiom of Choice and Zorn's Lemma

Any indexed family of sets $\mathscr{A} = \{A_i: i \in I\}$ may be conceived as a *variable set*, to wit, as a set *varying* over the index set *I.* Each A_i is then the "value" of the variable set $\mathscr{A}$ at *stage i.* A *choice function* on $\mathscr{A}$ is a map $f: I \to \bigcup_{i \in I} A_i$ such that $f(i) \in A_i$ for all $i \in I$. A choice function on $\mathscr{A}$ is thus a choice of an element of the variable set $\mathscr{A}$ at each stage; in other words, a choice function on $\mathscr{A}$ is just a *variable* (or global) *element* of $\mathscr{A}$. The *axiom of choice* (**AC**) asserts that if each member of a family $\mathscr{A}$ is nonempty, then there is a choice function on $\mathscr{A}$. Metaphorically, then, the axiom of choice asserts that any family of sets with an element at each stage has a variable or global element.

The corresponding "nonindexed" version of the axiom of choice asserts that if $\mathscr{B}$ is a family of nonempty sets then there is a map $f: \mathscr{B} \to \cup \mathscr{B}$—a *choice function* on $\mathscr{B}$—such that $f(X) \in X$ for each $X \in \mathscr{B}$.

The axiom of choice may also be formulated in the following way: corresponding to any family $\mathscr{B}$ of mutually disjoint nonempty sets there is a *choice set,* that is, a subset $C \subseteq \cup\, \mathscr{B}$ for which each intersection $C \cap B$ for $B \in \mathscr{B}$ has exactly one element. In this form the axiom of choice is sometimes provided with a "combinatorial" justification along the following lines. Given a family $\mathscr{B}$ of mutually disjoint nonempty sets, call a subset $S \subseteq \cup\, \mathscr{B}$ a *selector* for $\mathscr{B}$ if $S \cap B \neq \varnothing$ for all $B \in \mathscr{B}$. Clearly selectors for $\mathscr{B}$ exist; $\cup\, \mathscr{B}$ itself is an example. Now one can imagine starting with a selector S for $\mathscr{B}$ and "thinning out" each intersection $S \cap B$ for $B \in \mathscr{B}$ until it contains just a single element. The result is a choice set for $\mathscr{B}$. As is shown below, this argument, suitably refined, yields a precise derivation of the axiom of choice from Zorn's lemma.

Chronology of AC

1904/1908. Zermelo introduces axioms of set theory, explicitly formulates **AC** and uses it to prove the well-ordering theorem, thereby raising a storm of controversy.

1904. Russell recognizes **AC** as the *multiplicative axiom*: the product of arbitrary nonzero cardinal numbers is nonzero.

1914. Hausdorff derives from **AC** the existence of nonmeasurable sets in the "paradoxical" form that ½ of a sphere is congruent to 1/3 of it.

1922. Fraenkel establishes independence of **AC** from a system of set theory with atoms.

1924. Banach and Tarski derive from **AC** their paradoxical decompositions of the sphere: any solid sphere can be split into finitely many pieces which can be reassembled to form two solid spheres of the same size; and any solid sphere can be split into finitely many pieces in such a way as to enable them to be reassembled to form a solid sphere of arbitrary size.

1926. Hilbert introduces into his proof theory the "transfinite" or "epsilon" axiom as a version of **AC.**

1938. Gödel establishes relative consistency of **AC** and generalized continuum hypothesis with the axioms of set theory.

1963. Cohen proves independence of **AC** and continuum hypothesis from axioms of set theory.

Maximal Principles and Zorn's Lemma

A family of sets is *inductive* if it is closed under unions of chains. In its most typical form *Zorn's Lemma* **(ZL)** asserts that any nonempty inductive family of sets has a maximal element, that is, a member properly included in no member of the family. Zorn's lemma may also be formulated in a dual form. Call a family of sets *reductive* if it is closed under intersections of chains. Then any nonempty reductive family of sets has a minimal element, that is, a member properly including no member of the family.

Zorn's lemma and the axiom of choice are set-theoretically equivalent, but it is much more difficult to derive the former from the latter than vice-versa. There are two essentially different ways of deriving **AC** from **ZL**, a familiar one using maximal elements, and a less familiar one using minimal elements. The first derivation : if $\mathscr{A} = \{A_i: i \in I\}$ is an indexed family of nonempty sets, a *partial choice function* on $\mathscr{A}$ is a map f with domain $J \subseteq I$ such that $f(i) \in A_i$ for all $i \in J$. Now the maximal elements of the set $\mathscr{P}$ of partial choice functions on $\mathscr{A}$ are precisely the choice functions on $\mathscr{A}$; since $\mathscr{P}$ is clearly nonempty and readily shown to be inductive, Zorn's lemma yields the existence of a choice function on $\mathscr{A}$.

The second derivation of **AC** from **ZL** resembles the "combinatorial" justification of **AC** sketched above. Accordingly suppose given a family $\mathscr{B}$ of mutually disjoint nonempty sets; call a subset $S \subseteq \cup\mathscr{B}$ a *strong selector* for $\mathscr{B}$ if, for any $B \in \mathscr{B}$, either $B \subseteq S$ or $S \cap B$ is nonempty and finite. Now the *minimal* elements of the set $\mathscr{S}$ of strong selectors for $\mathscr{B}$ are precisely the choice sets for $\mathscr{B}$[1]; and $\mathscr{S}$ is clearly nonempty since it contains $\cup\mathscr{B}$. So if it can be shown that $\mathscr{S}$ is reductive[2], Zorn's lemma will yield a minimal element of $\mathscr{S}$, and so a choice set for $\mathscr{B}$. The reductiveness of $\mathscr{S}$ may be seen as follows: suppose that $\{S_i : i \in I\}$ is a chain of strong selectors; let $S = \bigcap_{i \in I} S_i$. We need to show that S is itself a strong selector; to this end let $B \in \mathscr{B}$ and suppose $B \not\subseteq S$. Then there is $i \in I$ for which $B \not\subseteq S_i$; since S_i is a strong selector, $S_i \cap B$ is finite nonempty, say $S_i \cap B = \{x_1, \ldots, x_n\}$. Clearly $S \cap B$ is then finite; suppose for the sake of contradiction that $S \cap B = \varnothing$. Then for each $k = 1, \ldots, n$ there is $i_k \in I$ for which $x_k \notin S_{i_k}$. It follows that $S_i \not\subseteq S_{i_k}$ for $k = 1, \ldots, n$, so, since the S_i form a chain, each S_{i_k} is a subset of S_i. Let S_j be

---

[1] That minimal strong selectors are choice sets is not entirely obvious. Suppose S is a minimal member of $\mathscr{S}$; then, given $B \in \mathscr{B}$, either (1) $S \cap B$ is finite nonempty or (2) $B \subseteq S$. In case (1) $S \cap B$ cannot contain two distinct elements because the removal of one of them from S would yield a strong selector smaller than S, violating its minimality. So in this case $S \cap B$ must be a singleton. In case (2) B cannot contain two distinct elements a, b since, if it did, $S' = [(S - B) \cup \{a\}]$ would be a strong selector smaller than S (notice that $S' \cap B = \{a\}$ and the relations of S' with the members of $\mathscr{S} - \{B\}$ are the same as those of S), again violating the minimality of S. So in this case B, and *a fortiori* $S \cap B$, must again be a singleton.

[2] Notice that, had we attempted to follow more closely the intuitive combinatorial derivation of **AC** as sketched above by using selectors instead of strong selectors we would have encountered the obstacle that—unlike the set of strong selectors—the set of selectors is not necessarily reductive.

the least of $S_{i_1},\ldots,S_{i_k}$; then $S_j \subseteq S_i$ But since $x_k \notin S_j$ for $k = 1, \ldots, n$, it now follows that $S_j \cap B = \varnothing$, contradicting the fact that S_j is a strong selector. Therefore $S \cap B \neq \varnothing$; and S is a strong selector as claimed.

Chronology of Maximal Principles

1909. Hausdorff introduces first explicit formulation of a maximal principle and derives it from **AC.**

1914. Hausdorff's *Grundzüge der Mengenlehre* (one of the first books on set theory and general topology) includes several maximal principles.

1922. Kuratowski formulates and employs several maximal principles in avoiding use of transfinite ordinals.

1926-28. S. Bochner and others independently introduce maximal principles.

1935. Max Zorn, unacquainted with previous formulations of maximal principles, publishes definitive version of maximal principle later to become celebrated as his lemma (**ZL). ZL** was first formulated in Hamburg in 1933, where Chevalley and Artin quickly "adopted" it. It seems to have been Artin who first recognized that **ZL** would yield **AC**, so that the two are equivalent (over the remaining axioms of set theory). Zorn regarded his principle less as a theorem than as an *axiom*—he hoped would it supersede cumbersome applications in algebra of transfinite induction and well-ordering, which algebraists in the Noether school had come to regard as "transcendental" devices.

1939-40. Teichmüller, Bourbaki and Tukey independently reformulate **ZL** in terms of "properties of finite character".

A Sampling of Applications of **AC—***all more elegantly proved via* **ZL**

Every vector space has a basis (initiated by Hamel in 1905). This was proved equivalent to **AC** by Blass 1984.

Every field has an algebraic closure (Steinitz 1910). This assertion is weaker than **AC**, indeed is a consequence of the (weaker) compactness theorem for first-order logic.

There is a Lebesgue nonmeasurable set of real numbers (Vitali 1905). This was shown much later to be a consequence of **BPI** (see below) and hence weaker than **AC.** Solovay 1965 established its independence of the remaining axioms of set theory

Every commutative ring with identity has a maximal ideal. This was proved equivalent to **AC** by Hodges 1979.

Every distributive lattice has a maximal ideal. Proved equibvalent to **AC** by Klimovsky 1958, and for lattices of sets by Bell and Fremlin 1972.

The Boolean Prime Ideal Theorem (BPI)—every Boolean algebra has a maximal (or prime) ideal. This was shown to be weaker than **AC** by Halpern and Levy 1966.

The Stone Representation Theorem for Boolean algebras (Stone 1936)—every Boolean algebra is isomorphic to a field of sets. This is equivalent to the **BPI** and hence weaker than **AC.**

The Sikorski Extension Theorem for Boolean algebras (Sikorski 1949)—every complete Boolean algebra is injective. The question of the equivalence of this with **AC** is one of the few remaining open questions in this area; it was proved independent of **BPI** by Bell 1983.

Every set is projective, or every surjection has a right inverse. Both of these are easily shown to be equivalent to **AC.**

Every free abelian group is projective. Proved equivalent to **AC** by Blass 1979.

Every divisible abelian group is injective. Proved equivalent to **AC** by Blass 1979.

The Hahn-Banach Theorem (1929)—given a sublinear functional *p* defined on a real linear space *X,* any *p*-dominated linear functional defined on a subspace of *X* can be extended to a *p*-dominated linear functional on *X.* This was later shown to be a consequence of **BPI** and hence weaker than **AC.**

Tychonov's Theorem (1930)—the product of compact spaces is compact. This was proved equivalent to **AC** by Kelley 1950. But for compact Hausdorff spaces it is equivalent to **BPI** (Rubin and Scott 1954) and hence weaker than **AC.**

Any continuous surjection between compact spaces has an irreducible restriction to a closed subset of its domain. (Here "irreducible" means: images of proper closed subsets are proper.) Proved equivalent to **AC** by Bell (1988), extending a result of Franklin and Thomas (1976).

Krein-Milman Theorem—the unit ball *B* of the dual of a real normed linear space has an extreme point, that is, one which is not an interior point of any line segment in *B*. This was proved equivalent to **AC** by Bell and Fremlin in 1972. There it is shown that, given any indexed family $\mathscr{A}$ of nonempty sets, there is a natural bijection between choice functions on $\mathscr{A}$ and the extreme points of the unit ball of the dual of a certain real normed linear space constructed from $\mathscr{A}$.

The compactness theorem for first-order logic (Gödel 1930, Malcev 1937, others)**—**if every finite subset of a of a set of first-order sentences has a model, then the set has a model. This was shown by Henkin in 1954 to be equivalent to **BPI**, and hence weaker than **AC.**

Completeness Theorem for First-Order Logic (Gödel 1930, Henkin 1954)—each consistent set of first-order sentences has a model. This was shown by Henkin in 1954 to be equivalent to **BPI**, and hence weaker than **AC.** If the cardinality of the model is specified in the right way, the assertion becomes equivalent to **AC.**

Löwenheim-Skolem-Tarski theorem (Löwenheim 1915, Skolem 1920, Tarski and Vaught 1957)—a first-order sentence having a model of cardinality κ also has a model of cardinality μ provided $\aleph_0 \leq \mu \leq \kappa$. This was proved equivalent to **AC** by Tarski.

AC implies the Law of Excluded Middle.

In 1975 Radu Diaconescu verified a conjecture of Bill Lawvere that any topos in which **AC** holds (in the version that every object is projective, or every epi has a right inverse) is Boolean, that is, every subobject of an object has a complement. This was later transformed into a very simple proof that **AC** implies the logical *law of excluded middle*: for any proposition $\boldsymbol{P}$, either $\boldsymbol{P}$ holds or its negation $\neg\boldsymbol{P}$ holds. (This may also be understood as asserting that the truth of any proposition is decidable.) In fact, to obtain this conclusion one only requires the axiom of choice for families of sets with just 2 elements, as the following argument shows:

Write 2 for the set 2 = {0, 1}. Let $\boldsymbol{P}$ be any proposition, and define

$$U = \{x \in 2\colon x = 0 \vee \boldsymbol{P}\} \qquad V = \{x \in 2\colon x = 1 \vee \boldsymbol{P}\}.$$

Now suppose given a choice function f on the family $\{U, V\}$. Writing $a = fU$, $b = fV$, we have $a \in U$, $b \in V$, i.e.,

$$[a = 0 \vee \boldsymbol{P}] \wedge [b = 1 \vee \boldsymbol{P}].$$

It follows that

$$[a = 0 \wedge b = 1] \vee \boldsymbol{P},$$

whence

(*) $$a \neq b \vee \boldsymbol{P},$$

Now clearly

$$\boldsymbol{P} \Rightarrow U = V = 2 \;\Rightarrow a = b,$$

whence

$$a \neq b \Rightarrow \neg\boldsymbol{P}.$$

But this and (*) together imply $\neg\boldsymbol{P} \vee \boldsymbol{P}$.