These remarks are intended as cultural background for Part II mathematicians.

In the wake of Russell's discovery of a contradiction in Frege's theory of classes, Ernst Zermelo, by training a physicist, turned his attention to the problem of axiomatising mathematics and proposed a system which with later enhancements by Adolf Fraenkel and Thoralf Skolem has become the standard set-theoretical basis of mathematics, and is known, unfairly to Skolem, as ZF.

Zermelo's idea was to escape from Russell's paradox by distinguishing two kinds of class: *sets*, which are classes that may be members of some class, and *proper classes*, which are not members of any class. So Russell's argument, which began "Consider the class of all classes that are not members of themselves" and reached a contradiction, now begins "Consider the class of all sets that are not members of themselves" and reaches only the harmless conclusion that the said class is not a set.

The Axiom of Choice

In 1904 Zermelo gave an explicit formulation of an additional principle that had surreptitiously been in use by mathematicians, much as Hilbert's axiomatisation of geometry made explicit principles of continuity used covertly by Euclid.

That principle is known as the Axiom of Choice, or AC for short. Its status relative to ZF was clarified in the course of the twentieth century:

Gödel (1935; announced in 1937): if ZF is consistent so is ZF plus AC
Cohen (1963): if ZF is consistent so is ZF plus the negation of AC.

Is AC true ?

In answering that, the analogy with the existence of Euclidean and of non-Euclidean geometries is to some extent helpful: the parallels postulate is neither provable nor refutable from the other Euclidean axioms, and the Axiom of Choice is neither provable nor refutable from the axioms of ZF. So if you ask "Should I believe AC or not ?", it is much as asking whether you should believe Euclid's parallels postulate. My retort in either case would be "Which world are you in ?" There are certain contexts where it is the right thing to assume; and there are others where it is not. Here are some examples.

AC in different branches of mathematics

In some branches of mathematics AC has found a huge number of applications; and many of these consequences—indeed several hundred—are known to be in fact equivalent to it. In algebra, the statement that every ring with an identity has a maximal ideal is equivalent to AC. In functional analysis, the conjunction of the Krein–Mil'man theorem and the Hahn–Banach theorem, both consequences of AC, implies AC. So to workers in those parts of mathematics, there are compelling reasons for accepting the axiom.

Number Theory gets the best of both worlds, for it is a consequence of Gödel's method for proving the consistency of AC relative to that of ZF that any arithmetical statement—for example, Fermat's Last Theorem—that has a proof in ZF + AC will have a proof in ZF alone. This "absoluteness argument" works also for certain statements of analysis; roughly those that are no more complicated than the statement that every differentiable function of a complex variable has a power series expansion at each point.

On the other hand, AC has consequences that are startling: the existence of sets of reals that are not Lebesgue measurable, and the famous Banach–Tarski paradoxical decomposition of a football into finitely many (non-measurable !) pieces that can be reassembled by translations and rotations to form two footballs of the same size as the original.

[I am a bit chary of describing anything as not intuitive, for one mathematician's strong intuition is another's blind spot; I think it would be fair to say that the Banach–Tarski paradox is shocking because one intuition has been brought into conflict with another. An argument originating in set theory (which I regard as abstract recursion theory and therefore a generalisation of arithmetic) has been imported into a geometrical context.]

A weak form of AC

So people working in, say, measure theory might well view AC with disfavour. One possibility is to work only with weak forms of AC: an interesting example is DC, the axiom of dependent choice, which avoids a possible pathology in the theory of well-founded relations, and which is more than enough to define Lebesgue measure.

[DC says that if a binary relation R on a set X is such that for each x in X there is a y in X with xRy, then there is an infinite sequence x_i $(i = 0, 1, 2, \ldots)$ of members of X such that for each i, $x_i R x_{i+1}$.]

An interesting way of negating AC

But other axioms have been proposed which contradict AC but have illuminating consequences. One such axiom has received much attention in recent years: the Axiom of Determinacy, AD, which guarantees the existence of winning strategies for one of the players in each of a natural family of games of infinite length between two players. AD supplies enough Choice to define Lebesgue measure, but not enough to build non-measurable sets; and indeed AD implies that all sets of reals are Lebesgue measurable.

The set-theorist's view

The favoured stance amongst set theorists is that AC is true in the whole universe of sets but that there are many interesting sub-universes (or inner models, as they are called) in which AC might well be false.

[Loosely, an inner model is a proper class M of sets with the property that members of members of M are members of M ("M is transitive"), and such that all the axioms of ZF are true when "set" is interpreted as "member of M". The study of relations between sub-universes resembles the study of relations between subfields of $\mathbb{C}$, though is as yet less developed.]

For example, there is a smallest inner model containing all the reals, called $L(\mathbb{R})$, and under relatively mild set-theoretical assumptions AD will be true in $L(\mathbb{R})$; and then many structure theorems become available for the study of Borel sets, analytic sets, and other subsets of $\mathbb{R}$ of interest to probabilists and chaos theorists.

On the other hand, in the smallest inner model of all, Gödel's constructible universe, L, AC is true in a very strong form, in that there is a single formula which defines a well-ordering of the whole of L; and the story does not stop there.

A shortcoming of Zorn's Lemma

The most common formulations of AC, are the original version, sometimes called the Multiplicative Principle, which states that the Cartesian product of a set of non-empty sets is non-empty; Zorn's Lemma, that a partial ordering, in which each chain has an upper bound, has a maximal element; and the Well-ordering Principle, that every set has a well-ordering.

Of those, Zorn's Lemma is perhaps the formulation most popular among mathematicians not interested in set theory or foundations. But it does have limitations. For set theorists the well-ordering principle is the preferred form because it links AC to abstract ideas of definition by recursion on well-founded relations. Once you start to look at the fine detail of a transfinite recursion, you are in a position to examine cases where AC does not suffice; and Zorn's Lemma tends to sweep such fine detail under the carpet.

Beyond AC

AC is of interest to mathematicians because it enables them to construct things. But AC is too weak for some constructions. The Continuum Hypothesis (which says that there are exactly as many real numbers as there are isomorphism types of countable well-orderings) is a good example of a principle that enables you to construct objects not otherwise available.

[Status: the generalised continuum hypothesis, GCH, implies AC (Sierpiński) and, of course, CH; GCH is true in L (Gödel).]

For example, in analysis a so-called p-point in the Stone–Cech compactification of the integers may be constructed using CH, (Walter Rudin, 1950) but it is known (Shelah, 1976) that $ZF + AC$ is consistent with the non-existence of any p-points.

For a more extreme example, in algebra, going beyond the power even of GCH to resolve, J.H.C.Whitehead asked whether every Abelian group G with $\mathrm{Ext}(G, \mathbb{Z})$ trivial is free Abelian. Whitehead's question was resolved by Shelah in 1974 when he proved that it is consistent with $ZF + AC$ for the answer to be negative, but that in L the answer is affirmative. He later proved that GCH is not strong enough for an affirmative answer.

The heart of this problem is that one wants to construct a free basis of a group, but there are many obstacles. Shelah used a principle formulated by Jensen, who showed it to be true in L, and known as the diamond principle, which supplies predictions of possible obstacles; there are more obstacles than predictions, but the predictions are sufficiently accurate to be used in the course of the construction to avoid the said obstacles. The interested reader can follow the story in *Almost Free Modules* by Eklof and Mekler.

There is no last word ...

So if you are doing Part II, feel free to use Zorn's Lemma, but remember that it is not the last word on the subject.

A. R. D. Mathias