5. The axiom of choice

We give a small number of equivalent forms of the axiom of choice; these forms should be sufficient for most mathematical purposes. The axiom of choice has been investigated a lot, and we give some references for this after proving the main theorem of this chapter.

The set of axioms of ZFC with the axiom of choice removed is denoted by ZF; so we work in ZF in this chapter.

The two main equivalents to the axiom of choice are as follows.

**Zorn’s Lemma.** If \((A, <)\) is a partial ordering such that \(A \neq \emptyset\) and every subset of \(A\) simply ordered by \(<\) has an upper bound, then \(A\) has a maximal element under \(<\), i.e., an element \(a\) such that there is no element \(b \in A\) such that \(a < b\).

**Well-ordering principle.** For every set \(A\) there is a well-ordering of \(A\), i.e., there is a relation \(<\) such that \((A, <)\) is a well-ordering.

In addition, the following principle, usually called the axiom of choice, is equivalent to the actual form that we have chosen:

**Choice-function principle.** If \(A\) is a family of nonempty sets, then there is a function \(f\) with domain \(A\) such that \(f(a) \in a\) for every \(a \in A\). Such a function \(f\) is called a choice function for \(A\).

**Theorem 5.1.** In ZF the following four statements are equivalent:

(i) the axiom of choice;

(ii) the choice-function principle;

(iii) Zorn’s lemma.

(iv) the well-ordering principle;

**Proof.** **Axiom of choice \(\Rightarrow\) choice-function principle.** Assume the axiom of choice, and let \(A\) be a family of nonempty sets. Let

\[ \mathcal{A} = \{ X : \exists a \in A[ X = \{(a, x) : x \in a]\} \}. \]

Since each member of \(A\) is nonempty, also each member of \(\mathcal{A}\) is nonempty. Given \(X, Y \in \mathcal{A}\) with \(X \neq Y\), choose \(a, b \in A\) such that \(X = \{(a, x) : x \in a\}\) and \(Y = \{(b, x) : x \in b\}\). Thus \(a \neq b\) since \(A \neq B\). The basic property of ordered pairs then implies that \(A \cap B = \emptyset\).

So, by the axiom of choice, let \(B\) have exactly one element in common with each element of \(\mathcal{A}\). Define \(f = \{ b \in B : \text{there exist } a \in A \text{ and } x \text{ such that } b = (a, x)\}\). Clearly \(f\) is the desired choice function for \(A\).

**Choice-function principle \(\Rightarrow\) Zorn’s lemma.** Assume the choice-function principle, and also assume the hypotheses of Zorn’s lemma. Let \(f\) be a choice function for \(\mathcal{P}(A) \setminus \{\emptyset\}\). We now define a function \(F : \text{Ord} \to A \cup \{A\}\) by recursion. Namely, for each ordinal \(\alpha\) let

\[
F(\alpha) = \begin{cases} f(\{a \in A : F(\beta) < a \text{ for all } \beta < \alpha\}) & \text{if this set is nonempty,} \\ A & \text{otherwise.} \end{cases}
\]

(1) If \(\alpha < \beta \in \text{Ord}\) and \(F(\beta) \neq A\), then \(F(\alpha) \neq A\), and \(F(\alpha) < F(\beta)\).
In fact, \( A \not\prec a \), so (1) is true by definition.

(2) There is an ordinal \( \alpha \) such that \( F(\alpha) = A \).

Otherwise, by (1), \( F \) is a one-one function from \( \text{Ord} \) into \( A \). So by the replacement axiom, \( \text{Ord} = F^{-1}[\text{rng}(F)] \) is a set, contradiction.

Let \( \alpha \) be minimum such that \( F(\alpha) = A \). By the hypothesis of Zorn’s lemma, \( \alpha \) is a successor ordinal \( \beta + 1 \). Thus \( F(\beta) \) is a \( \prec \)-maximal element of \( A \).

**Zorn’s lemma \( \Rightarrow \) well-ordering principle.** Assume Zorn’s lemma, and let \( A \) be any set. We may assume that \( A \) is nonempty. Let

\[
P = \{(B, \prec) : B \subseteq A \text{ and } (B \prec) \text{ is a well-ordering structure}\}.
\]

We partially order \( P \) as follows: \( (B, \prec) \prec (C, \ll) \iff B \subseteq C, \forall a, b \in B[a < b \iff a \ll b], \text{ and } \forall b \in B \forall c \in C \setminus B[b \ll c] \). Clearly this does partially order \( P \). \( P \neq \emptyset \), since \( (\{a\}, \emptyset) \in P \) for any \( a \in A \). Now suppose that \( Q \) is a nonempty subset of \( P \) simply ordered by \( \prec \). Let

\[
D = \bigcup_{(B, \prec) \in Q} B,
\]

\[
<_{D} = \bigcup_{(B, \prec) \in Q} <.
\]

Clearly \( (D, <_{D}) \) is a linear order with \( D \subseteq A \). Suppose that \( X \) is a nonempty subset of \( D \). Fix \( x \in X \), and choose \( (B, \prec) \in Q \) such that \( x \in B \). Then \( X \cap B \) is a nonempty subset of \( B \); let \( x \) be its least element under \( \prec \). Suppose that \( y \in X \) and \( y <_{D} x \). Choose \( (C, \ll) \in Q \) such that \( x, y \in C \) and \( y \ll x \). Since \( Q \) is simply ordered by \( \prec \), we have two cases.

**Case 1.** \( (C, \ll) \preceq (B, \prec) \). Then \( y \in C \subseteq B \) and \( y \in X \). so \( y < x \), contradicting the choice of \( x \).

**Case 2.** \( (B, \prec) \prec (C, \ll) \). If \( y \in B \), then \( y < x \), contradicting the choice of \( x \). So \( y \in C \setminus B \). But then \( x \ll y \), contradiction.

Thus we have shown that \( x \) is the \( <_{D} \)-least element of \( X \). So \( (D, <_{D}) \) is the desired upper bound for \( Q \).

Having verified the hypotheses of Zorn’s lemma, we get a maximal element \( (B, \prec) \) of \( P \). Suppose that \( B \neq A \). Choose any \( a \in A \setminus B \), and let

\[
C = B \cup \{a\},
\]

\[
<_{C} = < \cup \{(b, a) : b \in B\}.
\]

Clearly this gives an element \( (C, <_{C}) \) of \( P \) such that \( (B, \prec) \prec (C, <_{C}) \), contradiction.

**Well-ordering principle \( \Rightarrow \) Axiom of choice.** Assume the well-ordering principle, and let \( \mathcal{A} \) be a family of pairwise disjoint nonempty sets. Let \( C = \bigcup \mathcal{A} \), and let \( \prec \) be a well-order of \( C \). Define \( B = \{c \in C : c \text{ is the } \prec \text{-least element of the } P \in A \text{ for which } c \in P\} \). Clearly \( B \) has exactly one element in common with each member of \( A \).

There are many statements which are equivalent to the axiom of choice on the basis of ZF. We list some striking ones. A fairly complete list is in

37

(About 100 forms are listed, with proofs of equivalence.)

1. For every relation $R$ there is a function $f \subseteq R$ such that $\text{dmn}(f) = \text{dmn}(R)$.
2. For any sets $A, B$, either there is an injection of $A$ into $B$ or one of $B$ into $A$.
3. For any transitive relation $R$ there is a maximal $S \subseteq R$ which is a linear ordering.
4. Every product of compact spaces is compact.
5. Every formula having a model of size $\omega$ also has a model of any infinite size.
6. If $A$ can be well-ordered, then so can $\mathcal{P}(A)$.

There are also statements which follow from the axiom of choice but do not imply it on the basis of ZF. A fairly complete list of such statements is in


(383 forms are listed)

Again we list some striking ones:

1. Every Boolean algebra has a maximal ideal.
2. Any product of compact Hausdorff spaces is compact.
3. The compactness theorem of first-order logic.
4. Every commutative ring has a prime ideal.
5. Every set can be linearly ordered.
6. Every linear ordering has a cofinal well-ordered subset.
7. The Hahn-Banach theorem.
8. Every field has an algebraic closure.
9. Every family of unordered pairs has a choice function.
10. Every linearly ordered set can be well-ordered.

Now for the rest of this chapter we give one of the most startling consequences of the axiom of choice, the Banach-Tarski paradox. (This is optional material.)

The Banach-Tarski paradox is that a unit ball in Euclidean 3-space can be decomposed into finitely many parts which can then be reassembled to form two unit balls in Euclidean 3-space (maybe some parts are not used in these reassemblies). Reassembling is done using distance-preserving transformations. This is one of the most striking consequences of the axiom of choice, and is good background for the study of measure theory (of course the parts of the decomposition are not measurable). We give a proof of the theorem here
without going into any side issues. We follow Wagon, The Banach-Tarski paradox, where variations and connections to measure theory are explained in full. The proof involves very little set theory, only the axiom of choice. Some third semester calculus and some linear algebra and simple group theory are used. Altogether the proof should be accessible to a first-year graduate student who has seen some applications of the axiom of choice.

We start with some preliminaries on geometry and linear algebra. The “reassembling” mentioned in the Banach-Tarski paradox is entirely done by rotations and translations. Given a line in 3-space and an angle $\xi$, we imagine the rotation about the given line through the angle $\xi$. Mainly we will be interested in rotations about lines that go through the origin. We indicate how to obtain the matrix representations of such rotations. First suppose that $\varphi$ is the rotation about the $z$-axis counterclockwise through the angle $\xi$. Then, using polar coordinates,

$$
\varphi \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \varphi \begin{pmatrix} r \cos \theta \\ r \sin \theta \\ z \end{pmatrix} = \begin{pmatrix} r \cos(\theta + \xi) \\ r \sin(\theta + \xi) \\ z \end{pmatrix} = \begin{pmatrix} r \cos \theta \cos \xi - r \sin \theta \sin \xi \\ r \cos \theta \sin \xi + r \sin \theta \cos \xi \\ z \end{pmatrix} = \begin{pmatrix} x \cos \xi - y \sin \xi \\ x \sin \xi + y \cos \xi \\ z \end{pmatrix},
$$

which gives the matrix representation of $\varphi$:

$$
\begin{pmatrix} 
\cos \xi & -\sin \xi & 0 \\
\sin \xi & \cos \xi & 0 \\
0 & 0 & 1 
\end{pmatrix}.
$$

Similarly, the matrix representations of rotations counterclockwise through the angle $\xi$ about the $x$- and $y$-axes are, respectively,

$$
\begin{pmatrix} 
1 & 0 & 0 \\
0 & \cos \xi & -\sin \xi \\
0 & \sin \xi & \cos \xi 
\end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 
\cos \xi & 0 & \sin \xi \\
0 & 1 & 0 \\
-\sin \xi & 0 & \cos \xi 
\end{pmatrix}.
$$

Next, note that any rotation with respect to a line through the origin can be obtained as a composition of rotations about the three axes. This is easy to see using spherical coordinates. If $l$ is a line through the origin and a point $P$ different from the origin with spherical coordinates $\rho, \varphi, \theta$, a rotation about $l$ through an angle $\xi$ can be obtained as
follows: rotate about the $z$ axis through the angle $-\theta$, then about the $y$-axis through the angle $-\varphi$ (thereby transforming $l$ into the $z$-axis), then about the $z$-axis through the angle $\xi$, then back through $\varphi$ about the $y$ axis and through $\theta$ about the $z$-axis.

We want to connect this to linear algebra. Recall that a $3 \times 3$ matrix $A$ is orthogonal provided that it is invertible and $A^T = A^{-1}$. Thus the matrices above are orthogonal. A matrix is orthogonal iff its columns form a basis for $\mathbb{R}^3$ consisting of mutually orthogonal unit vectors; this is easy to see. It is easy to check that a product of orthogonal matrices is orthogonal. Hence all of the rotations about lines through the origin are represented by orthogonal matrices.

**Lemma 5.2.** If $A$ is an orthogonal $3 \times 3$ real matrix and $X$ and $Y$ are $3 \times 1$ column vectors, then $(AX) \cdot (AY) = X \cdot Y$, where $\cdot$ is scalar multiplication.

**Proof.** This is a simple computation:

\[(AX) \cdot (AY) = (AX)^T (AY) = X^T A^T A Y = X^T A^{-1} A Y = X^T Y = X \cdot Y.\]

It follows that any rotation about a line through the origin preserves distance, because $|P - Q| = \sqrt{(P - Q) \cdot (P - Q)}$ for any vectors $P$ and $Q$. Such rotations have an additional property: their matrix representations have determinant 1. This is clear from the discussion above. It turns out that this additional property characterizes the rotations about lines through the origin (see M. Artin, *Algebra*), but we do not need to prove that. The following property of such matrices is very useful, however.

**Lemma 5.3.** Suppose that $A$ is an orthogonal $3 \times 3$ real matrix with determinant 1, $A$ not the identity. Then there is a non-zero $3 \times 1$ matrix $X$ such that for any $3 \times 1$ matrix $Y$,

\[AY = Y \text{ iff } \exists a \in \mathbb{R} \left[ Y = aX \right].\]

**Proof.** Note that $A^T (A - I) = I - A^T = (I - A)^T$. Hence


It follows that $|A - I| = 0$. Hence the system of equations $(A - I)X = 0$ has a nontrivial solution, which gives the $X$ we want. Namely, we then have $AX = X$, of course. Then $A(aX) = aAX = aX$. This proves $\Leftarrow$ in the equivalence of the lemma. It remains to do the converse. We may assume that $X$ has length 1. Now we apply the Gram-Schmidt process to get a basis for $\mathbb{R}^3$ consisting of mutually orthogonal unit vectors, the first of which is $X$. We put them into a matrix $B$ as column vectors, $X$ the first column. Note that the first column of $AB$ is $X$, since $AX = X$, and hence the first column of $B^{-1}AB$ is $(1 \ 0 \ 0)^T$. Since $B^{-1}AB$ is an orthogonal matrix, it follows because its columns are mutually orthogonal that it has the form

\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & a & b \\
0 & c & d
\end{pmatrix}.
\]
Now suppose that $AY = Y$. Let $B^{-1}Y = (u \ e \ f)^T$. Then

(1) $e = f = 0$.

For, suppose that (1) fails. Now $B^{-1}ABB^{-1}Y = B^{-1}AY = B^{-1}Y$, while a direct computation using the above form of $B^{-1}AB$ yields $B^{-1}ABB^{-1}Y = (u \ ae + bf \ ce + df)^T$. So we get the two equations

$$
\begin{align*}
ae + bf &= e \\
 ce + df &= f
\end{align*}
$$

Since (1) fails, it follows that the determinant $\begin{vmatrix} a - 1 & b \\ c & d - 1 \end{vmatrix}$ is 0. Thus $ad - a - d + 1 - bc = 0$. Now $B^{-1}AB$ has determinant 1, and its determinant is $ad - bc$, so we infer that $a + d = 2$. But $a^2 + c^2 = 1$ and $b^2 + d^2 = 1$ since the columns of $B^{-1}AB$ are unit vectors, so $|a| \leq 1$ and $|d| \leq 1$. Hence $a = d = 1$ and $b = c = 0$. So $B^{-1}AB$ is the identity matrix, so $A$ is also, contradiction. Hence (1) holds after all.

From (1) we get $Y = B(u \ 0 \ 0)^T = uX$, as desired. \hfill \square

One more remark on geometry: any rotation preserves distance. We already said this for rotations about lines through the origin. If $l$ does not go through the origin, one can use a translation to transform it into a line through the origin, do the rotation then, and then translate back. Since translations clearly preserve distance, so arbitrary rotations preserve distance.

The first concrete step in the proof of the Banach, Tarski theorem is to describe a very special group of permutations of $3\mathbb{R}$. Let $\varphi$ be the counterclockwise rotation about the $z$-axis through the angle $\cos^{-1}(\frac{1}{3})$, and let $\rho$ the similar rotation about the $x$-axis. The matrix representation of these rotations and their inverses is, by the above,

$$
\varphi^{\pm 1} = \begin{pmatrix}
\frac{1}{3} \pm \frac{2\sqrt{2}}{3} & 0 \\
\frac{1}{3} & 0 \\
0 & 1
\end{pmatrix}, \quad \rho^{\pm 1} = \begin{pmatrix}
1 & 0 & 0 \\
0 & \frac{1}{3} & \pm \frac{2\sqrt{2}}{3} \\
0 & \frac{1}{3} & 1
\end{pmatrix}.
$$

Let $G_0$ be the group of permutations of $3\mathbb{R}$ generated by $\varphi$ and $\rho$. By a word in $\varphi$ and $\rho$ we mean a finite sequence with elements in $\{\varphi, \varphi^{-1}, \rho, \rho^{-1}\}$. Given such a word $w = \langle \sigma_0, \ldots, \sigma_{m-1} \rangle$, we let $\overline{w}$ be the composition $\sigma_0 \circ \ldots \circ \sigma_{m-1}$. Further, we call $w$ reduced if no two successive terms of $w$ have any of the four forms $\langle \varphi, \varphi^{-1} \rangle$, $\langle \varphi^{-1}, \varphi \rangle$, $\langle \rho, \rho^{-1} \rangle$, or $\langle \rho^{-1}, \rho \rangle$.

**Lemma 5.4.** If $w$ is a reduced word of positive length, then $\overline{w}$ is not the identity.

**Proof.** Suppose the contrary. Since $\rho \circ \overline{w} \circ \rho^{-1}$ is also the identity, we may assume that $w$ ends with $\rho^{\pm 1}$ (on the right). [If $w$ already ends with $\rho^{\pm 1}$, we do nothing. If it ends with $\varphi^{\pm 1}$, let $w' = \rho w \rho^{-1}$. Then $w'$ is reduced, unless $w$ has the form $\rho^{-1} w''$, in which case $w'' \rho^{-1}$ is reduced, and still $\overline{w''} \rho^{-1} = \overline{w} =$ the identity.]

Since obviously $\rho^{\pm 1}$ is not the identity, $w$ must have length at least 2. Now we claim
(1) For every terminal segment \( w' \) of \( w \) of nonzero even length the vector \( \overline{w'}(1 \ 0 \ 0)^T \) has the form \( (1/3^k)(a \ \ b\sqrt{2} \ c)^T \), with \( a \) divisible by 3 and \( b \) not divisible by 5.

We prove this by induction on the length of \( w' \). First note that, by computation,

\[
\begin{align*}
\rho \varphi &= \frac{1}{9} \begin{pmatrix} 3 & -6\sqrt{2} & 0 \\ 2\sqrt{2} & 1 & -6\sqrt{2} \\ 8 & 2\sqrt{2} & 3 \end{pmatrix} ; \\
\rho \varphi^{-1} &= \frac{1}{9} \begin{pmatrix} 3 & 6\sqrt{2} & 0 \\ -2\sqrt{2} & 1 & -6\sqrt{2} \\ -8 & 2\sqrt{2} & 3 \end{pmatrix} ; \\
\rho^{-1} \varphi &= \frac{1}{9} \begin{pmatrix} 3 & -6\sqrt{2} & 0 \\ 2\sqrt{2} & 1 & 6\sqrt{2} \\ -8 & -2\sqrt{2} & 3 \end{pmatrix} ; \\
\rho^{-1} \varphi^{-1} &= \frac{1}{9} \begin{pmatrix} 3 & 6\sqrt{2} & 0 \\ -2\sqrt{2} & 1 & 6\sqrt{2} \\ -8 & -2\sqrt{2} & 3 \end{pmatrix} .
\end{align*}
\]

Now we proceed by induction. For \( w' \) of length 2 we have

\[
\begin{align*}
\rho \varphi \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} &= \frac{1}{9} \begin{pmatrix} 3 \\ 2\sqrt{2} \\ 8 \end{pmatrix} ; \\
\rho \varphi^{-1} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} &= \frac{1}{9} \begin{pmatrix} 3 \\ -2\sqrt{2} \\ -8 \end{pmatrix} ; \\
\rho^{-1} \varphi \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} &= \frac{1}{9} \begin{pmatrix} 3 \\ 2\sqrt{2} \\ -8 \end{pmatrix} ; \\
\rho^{-1} \varphi^{-1} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} &= \frac{1}{9} \begin{pmatrix} 3 \\ -2\sqrt{2} \\ -8 \end{pmatrix} ;
\end{align*}
\]

hence (1) holds in this case. The induction step:

\[
\begin{align*}
\rho \varphi \begin{pmatrix} \frac{1}{3^k} \\ a \sqrt{2} \\ b \sqrt{2} \end{pmatrix} &= \frac{1}{3^{k+2}} \begin{pmatrix} 3a - 12b \\ 2\sqrt{2}\ a + b\sqrt{2} - 6\sqrt{2}c \\ 8a + 4b + 3c \end{pmatrix} ; \\
\rho \varphi^{-1} \begin{pmatrix} \frac{1}{3^k} \\ a \sqrt{2} \\ b \sqrt{2} \end{pmatrix} &= \frac{1}{3^{k+2}} \begin{pmatrix} 3a + 12b \\ -2\sqrt{2}\ a + b\sqrt{2} - 6\sqrt{2}c \\ -8a + 4b + 3c \end{pmatrix} ; \\
\rho^{-1} \varphi \begin{pmatrix} \frac{1}{3^k} \\ a \sqrt{2} \\ b \sqrt{2} \end{pmatrix} &= \frac{1}{3^{k+2}} \begin{pmatrix} 3a - 12b \\ 2\sqrt{2}\ a + b\sqrt{2} + 6\sqrt{2}c \\ -8a + 4b + 3c \end{pmatrix} ; \\
\rho^{-1} \varphi^{-1} \frac{1}{3^{k+2}} \begin{pmatrix} \frac{1}{3^k} \\ a \sqrt{2} \\ b \sqrt{2} \end{pmatrix} &= \left( \begin{pmatrix} 3a + 12b \\ -2\sqrt{2}\ a + b\sqrt{2} + 6\sqrt{2}c \\ 8a - 4b + 3c \end{pmatrix} \right) .
\end{align*}
\]

So, our assertion (1) is true. If \( w \) itself is of even length, then a contradiction has been reached, since \( b \) is not divisible by 5. If \( w \) is of odd length, then the following shows that the second entry of \( \overline{w} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}^T \) is nonzero, still a contradiction:

\[
\varphi \begin{pmatrix} \frac{1}{3^k} \\ a \sqrt{2} \\ b \sqrt{2} \end{pmatrix} = \frac{1}{3^{k+1}} \begin{pmatrix} a - 4b \\ 2\sqrt{2}\ a + b\sqrt{2} \\ 3c \end{pmatrix} .
\]
This finishes the proof of Lemma 5.4

This lemma really says that $G_0$ is (isomorphic to) the free group on two generators. But we do not need to go into that. We do need the following corollary, though.

**Corollary 5.5.** For every $g \in G_0$ there is a unique reduced word $w$ such that $g = \overline{w}$.

**Proof.** Suppose that $w$ and $w'$ both work, and $w \neq w'$. Say $w = \langle \sigma_0, \ldots, \sigma_{m-1} \rangle$ and $w' = \langle \tau_0, \ldots, \tau_{n-1} \rangle$. If one is a proper segment of the other, say by symmetry $w$ is a proper segment of $w'$, then

$$g = \overline{w} = \sigma_0 \circ \cdots \circ \sigma_{m-1}$$

$$= \overline{w'} = \tau_0 \circ \cdots \circ \tau_{n-1};$$

since $\sigma_i = \tau_i$ for all $i < m$, we obtain $I = \tau_m \circ \cdots \circ \tau_{n-1}$, $I$ the identity. But $\langle \tau_m, \ldots, \tau_{n-1} \rangle$ is reduced, contradicting 5.4.

Thus neither $w$ nor $w'$ is a proper initial segment of the other. Hence there is an $i < \min(m, n)$ such that $\sigma_i \neq \tau_i$, while $\sigma_j = \tau_j$ for all $j < i$ (maybe $i = 0$). But then we have by cancellation $\sigma_i \circ \cdots \circ \sigma_{m-1} = \tau_i \circ \cdots \circ \tau_{n-1}$, so $\tau_{n-1}^{-1} \circ \cdots \circ \tau_i^{-1} \sigma_i \circ \cdots \circ \sigma_{m-1} = I$. But since $\sigma_i \neq \tau_i$, the word $\langle \tau_{n-1}^{-1}, \ldots, \tau_i^{-1}, \sigma_i, \ldots, \sigma_{m-1} \rangle$ is reduced, again contradicting 5.4.

If $G$ is a group and $X$ is a set, we say that $G$ acts on $X$ if there is a homomorphism from $G$ into the group of all permutations of $X$. Usually this homomorphism will be denoted by $\gamma$, so that $\gamma$ is the permutation of $X$ corresponding to $g \in G$. (Most mathematicians don’t even use $\gamma$, using the same symbol for elements of the group and for the image under the homomorphism.) An important example is: any group $G$ acts on itself by left multiplication. Thus for any $g \in G$, $\gamma : G \to G$ is defined by $\gamma(h) = g \cdot h$, for all $h \in G$.

Let $G$ act on a set $X$, and let $E \subseteq X$. Then we say that $E$ is $G$-paradoxical if there are positive integers $m, n$ and pairwise disjoint subsets $A_0, \ldots, A_{m-1}, B_0, \ldots, B_{n-1}$ of $E$, and elements $\langle g_i : i < m \rangle$ and $\langle h_i : i < n \rangle$ of $G$ such that $E = \bigcup_{i \leq m} \gamma_i[A_i]$ and $E = \bigcup_{j < n} h_j[B_j]$. Note that this comes close to the Banach-Tarski formulation, except that the sets $X$ and $E$ are unspecified.

**Lemma 5.6.** $G_0$, acting on itself by left multiplication, is $G_0$-paradoxical.

**Proof.** If $\sigma$ is one of $\phi^{\pm 1}$, $\rho^{\pm 1}$, we denote by $W(\sigma)$ the set of all reduced words beginning on the left with $\sigma$, and $\overline{W}(\sigma) = \{ \overline{w} : w \in W(\sigma) \}$. Thus, obviously,$

$$G_0 = \{ I \} \cup \overline{W}(\phi) \cup \overline{W}(\phi^{-1}) \cup \overline{W}(\rho) \cup \overline{W}(\rho^{-1}),$$

where $I$ is the identity element of $G_0$. These five sets are pairwise disjoint by 5.5. Thus the lemma will be proved, with $m = n = 2$, by proving the following two statements:

1. $G_0 = \overline{W}(\phi) \cup \overline{\phi[W(\phi^{-1})]}$. 

43
To see this, suppose that \( g \in G_0 \) and \( g \notin \overline{\mathcal{W}}(\phi) \). Write \( g = \varpi \), \( w \) a reduced word. Then \( w \) does not start with \( \phi \). Hence \( \phi^{-1}w \) is still a reduced word, and \( g = \phi \circ \phi^{-1} \circ \overline{w} \in \hat{\phi}[\overline{\mathcal{W}}(\phi^{-1})] \), as desired.

(2) \( G_0 = \overline{\mathcal{W}}(\rho) \cup \hat{\rho}[\overline{\mathcal{W}}(\rho^{-1})] \).

The proof is just like for (1). \( \square \)

We need two more definitions, given that \( G \) acts on a set \( X \). For each \( x \in X \), the \( G \)-orbit of \( x \) is \( \{ \hat{g}(x) : g \in G \} \). The set of \( G \)-orbits forms a partition of \( X \). We say that \( G \) is without non-trivial fixed points if for every non-identity \( g \in G \) and every \( x \in X \), \( \hat{g}(x) \neq x \).

The following lemma is the place in the proof of the Banach-Tarski paradox where the axiom of choice is used. Don’t jump to the conclusion that the proof is almost over; our group \( G_0 \) above has non-trivial fixed points, and so does not satisfy the hypothesis of the lemma. Some trickery remains to be done even after this lemma. [For example, all points on the \( z \)-axis are fixed by \( \varphi \).]

**Lemma 5.7.** Suppose that \( G \) is \( G \)-paradoxical and acts on a set \( X \) without non-trivial fixed points. Then \( X \) is \( G \)-paradoxical.

**Proof.** Let

\[
A_0, \ldots, A_{m-1}, B_0, \ldots, B_{n-1}, g_0, \ldots, g_{m-1}, h_0, \ldots, h_{n-1}
\]

be as in the definition of paradoxical. By AC, let \( M \) be a subset of \( X \) having exactly one element in common with each \( G \)-orbit. Then we claim:

(1) \( \langle \hat{g}[M] : g \in G \rangle \) is a partition of \( X \).

First of all, obviously each set \( \hat{g}[M] \) is nonempty. Next, their union is \( X \), since for any \( x \in X \) there is a \( y \in M \) which is in the same \( G \)-orbit as \( x \), and this yields a \( g \in G \) such that \( x = \hat{g}(y) \) and hence \( x \in \hat{g}[M] \). Finally, if \( g \) and \( h \) are distinct elements of \( G \), then \( \hat{g}[M] \) and \( \hat{h}[M] \) are disjoint. In fact, otherwise let \( y \) be a common element. Say \( \hat{g}(x) = y \), \( x \in M \), and \( \hat{h}(z) = y \), \( z \in M \). Then clearly \( x \) and \( z \) are in the same \( G \)-orbit, so \( x = z \) since they are “both” in \( M \). Then \( (g^{-1} \cdot h)^{-1}(z) = z \) and \( g^{-1} \cdot h \) is not the identity, contradicting the no non-trivial fixed point assumption. So, (1) holds.

Now let \( A_i^* = \bigcup_{g \in A_i} \hat{g}[M] \) and \( B_j^* = \bigcup_{g \in B_j} \hat{g}[M] \), for all \( i < m \) and \( j < n \).

(2) \( A_i^* \cap A_k^* = \emptyset \) if \( i < k < m \).

In fact, suppose that \( x \in A_i^* \cap A_k^* \). Then we can choose \( g \in A_i \) and \( h \in A_k \) such that \( x \in \hat{g}[M] \cap \hat{h}[M] \). But \( g \neq h \) since \( A_i \cap A_k = \emptyset \), so this contradicts (1). Similarly the following two conditions hold:

(3) \( B_i^* \cap B_k^* = \emptyset \) if \( i < k < n \).
(4) \( A_i^* \cap B_j^* = \emptyset \) if \( i < m \) and \( j < n \).
(5) \( \bigcup_{i < m} \hat{g}_i[A_i^*] = X \).

For, let \( x \in X \). Say by (1) that \( x \in \hat{g}[M] \). Then by the choice of the \( A_i \)’s there is an \( i < m \) such that \( g \in \hat{g}_i[A_i] \). Say \( h \in A_i \) and \( g = \hat{g}_i(h) = g_i \cdot h \). Since \( x \in \hat{g}[M] \), say \( x = \hat{g}(m) \)
with \( m \in M \). Then \( x = (g_i \cdot h)(m) = \hat{g}_i(h(m)) \). Now \( \hat{h}(m) \in \hat{h}[M] \subseteq A_i^* \), so \( x \in \hat{g}_i[A_i^*] \), as desired in (5).

(6) \( \bigcup_{i<n} \hat{h}_i[B_i^*] = X \).

This is proved similarly. \( \square \)

Let \( S^2 = \{ x \in \mathbb{R}^3 : |x| = 1 \} \) be the usual unit sphere. Now we can prove the first paradoxical result leading to the Banach-Tarski paradox:

**Theorem 5.8.** (Hausdorff) *There is a countable \( D \subseteq S^2 \) such that \( S^2 \setminus D \) is \( G_0 \)-paradoxical.*

**Proof.** Let \( D \) be the set of all fixed points of non-identity elements of \( G_0 \). By 5.3, \( D \) is countable. Now we claim that if \( \sigma \in G_0 \), then \( \sigma[S^2 \setminus D] = S^2 \setminus D \). For, assume that \( x \in S^2 \setminus D \) and \( \sigma(x) \in D \). Say \( \tau \in G_0 \), \( \tau \) not the identity, and \( \tau(\sigma(x)) = \sigma(x) \). Then \( \sigma^{-1} \tau \sigma(x) = x \). Now \( \sigma^{-1} \circ \tau \circ \sigma \) is not the identity, since \( \tau \) isn’t, so \( x \in D \), contradiction. This proves that \( \sigma[S^2 \setminus D] \subseteq S^2 \setminus D \). This holds for any \( \sigma \in G_0 \), in particular for \( \sigma^{-1} \), and applying \( \sigma \) to that inclusion yields \( S^2 \setminus D \subseteq \sigma[S^2 \setminus D] \), so the desired equality holds.

Thus \( G_0 \) acts on \( S^2 \setminus D \) without non-trivial fixed points. So by 5.6 and 5.7, \( S^2 \setminus D \) is \( G_0 \)-paradoxical. \( \square \)

Let us see how far we have to go now. This theorem only looks at the sphere, not the ball. A countable subset is ignored. Since the sphere is uncountable, this makes the result close to what we want. But actually there is a countable subset of the sphere which is dense on it. [Take points whose spherical coordinates are rational.]

For the next step we need a new notion. Suppose that \( G \) is a group acting on a set \( X \), and \( A, B \subseteq X \). We say that \( A \) and \( B \) are finitely \( G \)-equidecomposable if \( A \) and \( B \) can be decomposed into the same number of parts, each part of \( A \) being carried into the corresponding part of \( B \) by an element of \( G \). In symbols, there is a positive integer \( n \) such that there are partitions \( A = \bigcup_{i<n} A_i \) and \( B = \bigcup_{i<n} B_i \) and members \( g_i \in G \) for \( i < n \) such that \( g_i[A_i] = B_i \) for all \( i < n \). We then write \( A \sim_G B \).

**Lemma 5.9.** *If \( G \) acts on a set \( X \), then \( \sim_G \) is an equivalence relation on \( \mathcal{P}(X) \).*

**Proof.** Obviously \( \sim_G \) is reflexive on \( \mathcal{P}(X) \) and is symmetric. Now suppose that \( A \sim_G B \sim_G C \). Then we get partitions \( A = \bigcup_{i<m} A_i \) and \( B = \bigcup_{i<m} B_i \) with elements \( g_i \in G \) such that \( g_i[A_i] = B_i \) for all \( i < m \); and partitions \( B = \bigcup_{j<n} B'_j \) and \( C = \bigcup_{j<n} C_j \) with elements \( h_j \in G \) such that \( h_j[B'_j] = C_j \) for all \( j < n \). Now for all \( i < m \) and \( j < n \) let \( B_{ij} = B_i \cap B'_j \), \( A_{ij} = g_i^{-1}[B_{ij}] \), and \( C_{ij} = h_j[B_{ij}] \). Then \( A = \bigcup_{i<m,j<n} A_{ij} \) is a partition of \( A \), \( C = \bigcup_{i<m,j<n} C_{ij} \) is a partition of \( C \), and \( (h_j \cdot g_i)[A_{ij}] = C_{ij} \). Some of the \( B_{ij} \) may be empty; eliminating the empty ones yields the desired nonemptiness of members of the partitions. \( \square \)

**Lemma 5.10.** *Suppose that \( G \) acts on \( X \), \( E \) and \( E' \) are finitely \( G \)-equidecomposable subsets of \( X \), and \( E \) is \( G \)-paradoxical. Then also \( E' \) is \( G \)-paradoxical.*

**Proof.** Because \( E \) is \( G \)-paradoxical, we can find pairwise disjoint subsets

\[
A_0, \ldots, A_{m-1}, B_0, \ldots, B_{n-1}
\]
of $E$ and corresponding elements $g_0, \ldots, g_{m-1}, h_0, \ldots, h_{n-1}$ of $G$ such that

$$E = \bigcup_{i<m} \tilde{g}_i[A_i] = \bigcup_{j<n} \tilde{h}_j[B_i].$$

And because $E$ and $E'$ are finitely $G$-equidecomposable we can find partitions $E = \bigcup_{k<p} E_k$ and $E' = \bigcup_{k<p} E_k'$ with elements $\tilde{f}_i \in G$ such that $\tilde{f}_i[E_k] = E_k$ for all $k < p$. Then the following sets are pairwise disjoint: $A_i \cap \tilde{g}_i^{-1}[C_k]$ for $i < m$ and $k < p$, and $B_j \cap \tilde{h}_j^{-1}[C_k]$ for $j < n$ and $k < p$. And

$$E' = \bigcup_{k<p} D_k = \bigcup_{k<p} \tilde{f}_k[C_k]$$

$$= \bigcup_{k<p} \tilde{f}_k[C_k] \cap \bigcup_{i<m} \tilde{g}_i[A_i]$$

$$= \bigcup_{k<p, i<m} \tilde{f}_k[C_k] \cap \tilde{g}_i[A_i]$$

$$= \bigcup_{k<p, i<m} (f_k \cdot g_i)^{-1}[A_i \cap \tilde{g}_i^{-1}[C_k]],$$

and similarly

$$E' = \bigcup_{k<p, j<n} (f_k \cdot h_j)^{-1}[B_i \cap \tilde{h}_j^{-1}[C_k]].$$

\[ \square \]

**Lemma 5.11.** Let $D$ be a countable subset of $S^2$. Then there is a rotation $\sigma$ with respect to a line through the origin such that if $G_1$ is the group of transformations of $^3\mathbb{R}$ generated by $\sigma$, then $S^2$ and $S^2 \setminus D$ are $G_1$-equidecomposable.

**Proof.** For each $d \in D$ let $f(d)$ be the line through the origin and $d$. Then $f$ maps $D$ into the set $L$ of all lines through the origin, and the range of $f$ is countable. But $L$ itself is uncountable: for example, for each $\theta \in [0, \pi]$ one can take the line through the origin and $(\cos \theta, \sin \theta, 0)$. Hence there is a line $l \in L$ not in the range of $f$. This means that $l$ does not pass through any point of $D$. Fix a direction in which to take rotations about $l$.

Note that if $P$ and $Q$ are distinct points of $D$, then there is at most one rotation about $l$ which takes $P$ to $Q$ and is between 0 and $2\pi$; this will be denoted by $\psi_{PQ}$, if it exists. Now let $A$ consist of all $\theta \in (0, 2\pi)$ such that there is a positive integer $n$ and a $P \in D$ such that $\sigma$ is the rotation about $l$ through the angle $n\theta$, then $\sigma(P) \in D$. We claim that $A$ is countable. For, if $P, Q \in D$, $\psi_{PQ}$ is defined, $n \in \omega \setminus \{0\}$, $k \in \omega$, and $0 < \frac{1}{n}(\psi_{PQ} + 2\pi k) < 2\pi$, then $\frac{1}{n}(\psi_{PQ} + 2\pi k) \in A$; and every member of $A$ can be obtained this way. [Given $\theta \in A$, we have $n\theta = \psi_{PQ} + 2\pi k$ for some $P, Q \in D$ and $n, k \in \omega$.] This really defines a function from $D \times D \times (\omega \setminus \{0\}) \times \omega$ onto $A$, so $A$ is, indeed, countable. We choose $\theta \in (0, 2\pi) \setminus A$, and take the rotation $\sigma$ about $l$ through the angle $\theta$. Let $D = \bigcup_{n \in \omega} \sigma^n[D]$. The choice of $\sigma$ says that $\sigma^n[D] \cap D = 0$ for every positive integer $n$. Hence if $n < m < \omega$ we have $\sigma^n[D] \cap \sigma^m[D] = 0$, since

$$\sigma^n[D] \cap \sigma^m[D] = \sigma^n[D \cap \sigma^{m-n}[D]] = \sigma^n[0] = 0.$$
Note that \( \sigma[\mathcal{D}] = \mathcal{D} \setminus D \). Hence

\[
S^2 = \mathcal{D} \cup (S^2 \setminus \mathcal{D}) \sim_{G_1} \sigma[\mathcal{D}] \cup (S^2 \setminus \mathcal{D}) = S^2 \setminus D.
\]

Now let \( G_2 \) be the group of permutations of \( \mathbb{R}^3 \) generated by \( \{ \varphi, \rho, \sigma \} \). We now have the first form of the Banach-Tarski paradox:

**Theorem 5.12.** (Banach, Tarski) \( S^2 \) is \( G_2 \)-paradoxical.

One can loosely state this theorem as follows: one can decompose \( S^2 \) into a finite number of pieces, rotate some of these pieces finitely many times with respect to certain lines through the origin to reassemble \( S^2 \), and then similarly transform some of the remaining pieces to also reassemble \( S^2 \). The rotations are of three kinds: the very specific rotations \( \varphi \) and \( \rho \) defined at the beginning of this section, and the rotation \( \sigma \) in the preceding proof, for which we do not have an explicit description. One can apply the inverses of these rotations as well. After doing the second reassembling, one can apply a translation to make that copy of \( S^2 \) disjoint from the first copy.

Finally, let \( B = \{ x \in \mathbb{R}^3 : |x| \leq 1 \} \) be the unit ball in 3-space. Let \( G_3 \) be the group generated by \( \varphi, \rho, \sigma \), and the rotation \( \tau \) about the line determined by \((0, 0, \frac{1}{2})\) and \((1, 0, \frac{1}{2})\), through the angle \( \pi/\sqrt{2} \). Note that by the considerations at the beginning of this section, \( \tau \) consists of the translation \((x \ y \ z)^T \mapsto (x \ y \ z - \frac{1}{2})^T\), followed by the rotation through \( \frac{\pi}{\sqrt{2}} \) about the \( x \)-axis, followed by the translation \((x \ y \ z)^T \mapsto (x \ y \ z + \frac{1}{2})^T\).

**Lemma 5.13.** For any positive integer \( k \),

\[
\tau^k \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \sin \left( \frac{k\pi}{\sqrt{2}} \right) \\ \frac{1}{2} \cos \left( \frac{k\pi}{\sqrt{2}} \right) + \frac{1}{2} \end{pmatrix}.
\]

Hence \( \tau^p (0 \ 0 \ 0)^T \neq (0 \ 0 \ 0)^T \) for every positive integer \( p \).

**Proof.** The first equation is easily proved by induction on \( k \). Then the second inequality follows since for any positive integer \( p \), the argument \( \frac{p\pi}{\sqrt{2}} \) is never equal to \( m\pi \) for any integer \( m \), since \( \frac{\pi}{\sqrt{2}} \) is irrational.

**Theorem 5.14.** (Banach, Tarski) \( B \) is \( G_3 \)-paradoxical.

**Proof.** By 5.12 there are pairwise disjoint subsets \( A_i \) and \( B_j \) of \( S^2 \) and members \( g_i, h_j \) of \( G_2 \) for \( i < m \) and \( j < n \) such that \( S^2 = \bigcup_{i<m} g_i[A_i] = \bigcup_{j<n} h_j[B_j] \). For each \( i < m \) and \( j < n \) let \( A'_i = \{ \alpha x : x \in A_i, \ 0 < \alpha \leq 1 \} \) and \( B_j = \{ \alpha x : x \in B_j, \ 0 < \alpha \leq 1 \} \). Then

1. The \( A'_i \)'s and \( B'_j \)'s are pairwise disjoint.
2. \( B \setminus \{0\} = \bigcup_{i<m} g_i[A'_i] = \bigcup_{j<n} h_j[B'_j] \).
In fact, let \( y \in B \setminus \{0\} \). Let \( x = y/|y| \). Then \( x \in S^2 \), so there is an \( i < m \) such that \( x \in g_i[A_i] \). Say that \( x = g_i(z) \) with \( z \in A_i \). Then \( |y|z \in A'_i \), and \( g_i(|y|z) = |y|g_i(z) = |y|x = y \). So \( y \in g_i[A_i'] \). This proves the first equality in (2), and the second equality is proved similarly.

So far, we have shown that \( B \setminus \{0\} \) is \( G_2 \)-paradoxical. Now we show that \( B \) and \( B \setminus \{0\} \) are finitely \( G_3 \)-equidecomposable, which will finish the proof. By lemma 5.13 we have

\[
B = D \cup (B \setminus D) \sim_{G_3} \tau[D] \cup (B \setminus D) = B \setminus \{0\}
\]

This proves the desired equidecomposability.

As in the case of \( S^2 \), a translation can be made if one wants one copy of \( B \) to be disjoint from the other.

**EXERCISES**

In the first four exercises, we assume elementary background and ask for the proofs of some standard mathematical facts that require the axiom of choice.

E5.1. Show that any vector space over a field has a basis (possibly infinite).

E5.2. A subset \( C \) of \( \mathbb{R} \) is closed iff the following condition holds:

For every sequence \( f \in \mathbb{R}^\omega \), if \( f \) converges to a real number \( x \), then \( x \in C \).

Here to say that \( f \) converges to \( x \) means that

\[
\forall \varepsilon > 0 \exists M \forall m \geq M \exists f_m \in \mathbb{R}^\omega \text{ such that } |f_m - x| < \varepsilon.
\]

Prove that if \( \langle C_m : m \in \omega \rangle \) is a sequence of nonempty closed subsets of \( \mathbb{R} \), \( \forall m \in \omega \forall x, y \in C_m \exists f_m \in \mathbb{R}^\omega \text{ such that } |f_m - x| < 1/(m+1) \), and \( C_m \supseteq C_n \) for \( m < n \), then \( \bigcap_{m \in \omega} C_m \) is nonempty. Hint: use the Cauchy convergence criterion.

E5.3. Prove that every nontrivial commutative ring with identity has a maximal ideal. Nontrivial means that \( 0 \neq 1 \). Only very elementary definitions and facts are needed here; they can be found in most abstract algebra books. Hint: use Zorn’s lemma.

E5.4. A function \( g : \mathbb{R} \to \mathbb{R} \) is continuous at \( a \in \mathbb{R} \) iff for every sequence \( f \in \mathbb{R}^\omega \) which converges to \( a \), the sequence \( g \circ f \) converges to \( g(a) \). (See Exercise E5.2.) Show that \( g \) is continuous at \( a \) iff the following condition holds:

\[
\forall \varepsilon > 0 \exists \delta > 0 \forall x \in \mathbb{R} \exists f_m \in \mathbb{R}^\omega \exists f : |f(x) - a| < \delta \rightarrow |g(x) - g(a)| < \varepsilon.
\]

Hint: for \( \rightarrow \), argue by contradiction.

E5.5 Show by induction on \( m \), without using the axiom of choice, that if \( m \in \omega \) and \( \langle A_i : i \in m \rangle \) is a system of nonempty sets, then there is a function \( f \) with domain \( m \) such that \( f(i) \in A_i \) for all \( i \in m \).
E5.6 Using AC, prove the following, which is called the Principle of Dependent Choice (which is also weaker than the axiom of choice, but cannot be proved in ZF). If \( A \) is a nonempty set, \( R \) is a relation, \( R \subseteq A \times A \), and for every \( a \in A \) there is a \( b \in A \) such that \( aRb \), then there is a function \( f: \omega \to A \) such that \( f(i)Rf(i+1) \) for all \( i \in \omega \).

The remaining exercises outline proofs of some equivalents to the axiom of choice; so each exercise states something provable in ZF. We are interested in the following statements.

1. If \( < \) is a partial ordering and \( \preceq \) is a simple ordering which is a subset of \( < \), then there is a maximal (under \( \subseteq \) ) simple ordering \( \ll \) such that \( \preceq \) is a subset of \( \ll \), which in turn is a subset of \( < \).

2. For any two sets \( A \) and \( B \), either there is a one-one function mapping \( A \) into \( B \) or there is a one-one function mapping \( B \) into \( A \).

3. For any two nonempty sets \( A \) and \( B \), either there is a function mapping \( A \) onto \( B \) or there is a function mapping \( B \) onto \( A \).

4. A family \( \mathcal{F} \) of subsets of a set \( A \) has finite character if for all \( X \subseteq A \), \( X \in \mathcal{F} \) iff every finite subset of \( X \) is in \( \mathcal{F} \). Principle (4) says that every family of finite character has a maximal element under \( \subseteq \).

5. For any relation \( R \) there is a function \( f \subseteq R \) such that \( \text{dmn } R = \text{dmn } f \).

E5.7. Show that the axiom of choice implies (1). [Use Zorn’s lemma]

E5.8. Prove that (1) implies (2). [Given sets \( A \) and \( B \), define \( f < g \) iff \( f \) and \( g \) are one-one functions which are subsets of \( A \times B \), and \( f \subset g \). Apply (1) to \( < \) and the empty simple ordering.]

E5.9. Prove that (2) implies (3). [Easy]

E5.10. Show in ZF that for any set \( A \) there is an ordinal \( \alpha \) such that there is no one-one function mapping \( \alpha \) into \( A \). Hint: consider all well-orderings contained in \( A \times A \).

E5.11. Prove that (3) implies the axiom of choice. [Show that any set \( A \) can be well-ordered, as follows. Use exercise 4 to find an ordinal which cannot be mapped one-one into \( \mathcal{P}(A) \). Show that if \( f: A \to \alpha \) maps onto \( \alpha \), then \( \langle f^{-1}([\beta]) : \beta < \alpha \rangle \) is a one-one function from \( \alpha \) into \( \mathcal{P}(A) \).

E5.12. Show that the axiom of choice implies (4). [Use Zorn’s lemma.]

E5.13. Show that (4) implies (5). [Given a relation \( R \), let \( \mathcal{F} \) consist of all functions contained in \( R \).]

E5.14. Show that (5) implies the axiom of choice. [Given a family \( \langle A_i : i \in I \rangle \) of nonempty sets, let \( R = \{(i,x) : i \in I \text{ and } x \in A_i\} \).]