1 Sets

What are sets for?

We speak of collections all the time: this university consists of its colleges, faculties, etc; this lecture handout consists of its pages; the philosophy faculty library consists of its books. We speak of collections, aggregates, classes, sets. This part of the course will first be concerned with what these sets are. As well as being an important philosophical notion in itself, what purposes have sets been put to?

The first motivation is mathematical. At the end of the 17th century, Newton and Leibniz discovered the calculus (analysis). This is the study of real-valued functions of a real variable. Real numbers are rather different to the other sorts of number – they are infinite objects. And, in its original form, the calculus made use of the notion of ‘infinitesimal’. Generally, the notion of the infinite was becoming more central to mathematics. One crucial use of set theory is in taming the infinite.

Set theory has also come to serve a crucial, foundational role for mathematics. We would very much like our mathematical theories – of arithmetic, geometry, and so on – to be consistent. Virtually every mathematical theory of interest can be shown to be consistent if set theory is. So the consistency of set theory is of considerable philosophical interest.

There are other, more mundane, motivations. First, sets can provide a neat way of discussing semantics and metalogic. For example, what does the predicate ‘is red’ refer to? We might think redness, but what sort of object could that be? Rather, some have preferred to say that it refers to the set of red things. Similarly for metalogic: we might want to discuss the soundness and completeness of a proof system for a logic. How can we do this? One helpful way is to compare the set of logical truths to the set of theorems and ask if they are the same.
What are sets?

What is a set? It is a collection of objects, considered as a single object. Intuitively, you can think of a set as a shopping bag into which you can put all sorts of things. These can be concrete things like fruit or philosophers. Or they can be abstract entities like the natural numbers 0 to 10. They can be completely unrelated like the number 2, the concept of humility and The McConaissance.

The objects making up the set are called elements or members of the set. If \( o \) is a member of the set \( S \), we write \( o \in S \).

If \( a \) and \( b \) are the only members of a set, we can write the set as \( \{a, b\} \). The order of the objects in a set is insignificant: \( \{a, b\} = \{b, a\} \).

Sets can also be members of sets. So the set \( \{\{a, b\}, c\} \) is a set. It has two members: the object \( c \) and the set \( \{a, b\} \). The objects \( a \) and \( b \) are members of \( \{a, b\} \): \( a \in \{a, b\} \) and \( b \in \{a, b\} \). But they are not members of \( \{\{a, b\}, c\} \): \( a \notin \{\{a, b\}, c\} \) and \( b \notin \{\{a, b\}, c\} \).

So far, we have listed the members of a set when specifying it, but often this will be laborious or even impossible. Consider the set of philosophers. There are loads of us and we don’t want to have to list Frege, Russell, Wittgenstein, etc. Rather, we allow ourselves to specify a condition: \( \{x : x \text{ is a philosopher}\} \). Sometimes, the task of listing all of the members is impossible, because infinite, and we must use this method of notation: \( \{x : x \text{ is a natural number}\} \).

To be clear, the notation will often be a matter of convenience: \( \{2\} = \{x : x \text{ is an even prime number}\} \). These are two ways of naming the very same set. Nothing hinges on which one we choose. Generally, we’ll write \( \{x : \phi(x)\} \) for the set of \( \phi \)s.

The most basic principle governing the nature of sets is their Extensionality:

**Extensionality** Where \( A \) and \( B \) are sets: \( \forall x (x \in A \leftrightarrow x \in B) \rightarrow A = B \)

(Here and throughout, as is standard, I will use lower- and upper-case letters to stand for objects. The only motivation is readability.)

Extensionality delivers that, for any objects, there is exactly one set of which they are all members. We have seen that we don’t care about the order of the members of a set; we also don’t care about repetition: \( \{a, a, b\} = \{a, b\} = \{b, a\} \). In each case, we have named the same set: the set that has \( a \) and \( b \) as its only members.

Extensionality also gives us a means of proving that two sets are identical. We must establish two conditionals:

1. \( \forall x (x \in A \rightarrow x \in B) \)
2. \( \forall x (x \in B \rightarrow x \in A) \)
1. and 2. allow us to conclude that some object \( x \) is a member of \( A \) if, and only if, it is in \( B \). By Extensionality, we can then conclude that \( A = B \).

What are sets not? They are not mereological fusions of their members. The Lecture Block can be considered as the fusion of its parts, but a fusion is no more than the sum of its parts. The set of rooms in the Lecture Block is something over and above its members. Sets are also not the sums of physically scattered objects in the way that the United States might be regarded as the sum of its mainland, Alaska and Hawaii.

**Union**

We are now in a position to introduce two further set-theoretic notions: union and intersection. If \( A \) and \( B \) are sets, there is a set \( \{ x : x \in A \lor x \in B \} \). This is the set of all objects in either \( A \) or \( B \). Venn diagrams provide a helpful way of thinking about this:

![Venn Diagram for Union](image)

We will give this operation on sets a name and symbol:

**Union** The union of two sets \( A \) and \( B \), \( A \cup B \), is the set of all things which are members of \( A, B \) or both: \( A \cup B = \{ x : x \in A \lor x \in B \} \).

If \( A \) is the set of cats and \( B \) of dogs, then \( A \cup B \) is the set of things that are either cats or dogs. Continuing the shopping bag analogy, we can imagine tipping the contents of bags \( A \) and \( B \) into one larger bag. Consider \( \{0, 1, 2\} \) and \( \{2, 3, 4\} \). 
\( \{0, 1, 2\} \cup \{2, 3, 4\} = \{0, 1, 2, 2, 3, 4\} = \{0, 1, 2, 3, 4\} \). Union delivers the first identity; Extensionality the second. There is an obvious analogy between union and (inclusive) disjunction.

**Intersection**

Similarly, we can define the set-theoretic operation of intersection, which can be depicted as:

![Venn Diagram for Intersection](image)
The definition here is:

**Intersection** The *intersection* of two sets $A$ and $B$, $A \cap B$, is the set of all things that are members of *both* $A$ and $B$: $A \cap B = \{x : x \in A \land x \in B\}$.

For example, if $A$ is the set of all animals and $C$ is the set of all chimpanzees, the set $A \cap C$ is the set of all things that are both animals and chimpanzees, namely the set of all chimpanzees. If $C$ is the set of sitcom creators and $G$ of Grammy-winning musicians, $C \cap G$ is the set of all sitcom creators who are *also* Grammy-winning musicians, and Donald Glover $\in C \cap G$. There is an obvious analogy between intersection and conjunction.

**The empty set and singletons**

If $C$ is the set of all cats and $D$ of all dogs, the intersection $C \cap D$ has no members. We call $C$ and $D$ *disjoint*. When two sets $A$ and $B$ are disjoint, $A \cap B$ is empty.

By Extensionality, there is exactly one *empty set*, $\emptyset$. The notion of an empty set may sound odd but, continuing with the shopping bag analogy, this is analogous to a shopping bag with nothing in it.

For any object, there is a set – called its *singleton* – that contains only that set. The singleton of Donald Glover is $\{ \text{Donald Glover} \}$. Importantly, Donald Glover $\neq \{ \text{Donald Glover} \}$. Why? By Leibniz’s Law, since Donald Glover writes sitcoms and sets can’t write sitcoms!

Sets also have singletons, so the singleton of $\{ \text{Donald Glover} \}$ is $\{ \{ \text{Donald Glover} \} \}$. Again, $\{ \text{Donald Glover} \} \neq \{ \{ \text{Donald Glover} \} \}$. Why? They don’t have the same members: Donald Glover is a member of the former but not of the latter. By Extensionality, therefore, these are different sets. Again, nothing odd is going on here: continuing the shopping bag analogy, a singleton is a shopping bag with just one thing in it. We could then put this bag inside another, and we’d end up with a different bag, extensionally speaking.

**Subsets**

Another crucial notion is *subsethood*. It is *absolutely crucial* that subsethood is not confused with membership. Membership is the relation that holds between an *object* (possibly a set) and a set. Subsethood is a relation that holds between two sets. It holds between sets $A$ and $B$ just when every member of $A$ is also a member of $B$:

**Subset** If every member of set $A$ is also a member of $B$, then $A$ is a *subset* of $B$, $A \subseteq B$: $A \subseteq B \leftrightarrow \forall x(x \in A \rightarrow x \in B)$.

For example, the set of footballers $F$ is a subset of the set of sportspeople $S$, since every footballer is also a sportsperson: $F \subseteq S$. Of course, there are also
sportspeople who are not footballers, such as Ronnie O’Sullivan. In this case, 
$F \subseteq S$ and $F \neq S$. When this is the case, we say that the first is a proper 
subset of the second: $F \subset S$. Here there is an obvious notational analogy with e.g. ‘≤’.

Consider the set of all triangles $A$ and the set of all trilateral shapes $L$. Here, 
$A \subseteq L$ but of course $L \subseteq A$. When this holds, we can conclude $A = L$. This is of 
course what we’d expect. We’ve said that one test for whether two sets $A$ and $B$ 
are identical is whether it is the case that every member of $A$ is a member of $B$ 
and vice versa. We now know this equivalent to saying that $A \subseteq B$ and $B \subseteq A$.

Here are some examples to illustrate the differences between membership and 
subsethood, based on the sets $A = \{0, 1\}$ and $B = \{0, 1, 2, 3\}$.

- $A \subseteq B$ – this is true, because every member of $A$ is a member of $B$
- $A \subset B$ – this is also true because $A \subseteq B$ and $A \neq B$
- $A \in B$ – this is false because, although $A \subseteq B$, the set $\{0, 1\}$ is not a 
  member of $B$; indeed, no set is a member of $B$
- $0 \in A$ – this is true, because 0 is a member of $\{0, 1\}$
- $0 \subseteq B$ – this is ill-formed: ‘⊆’ must be flanked by names of sets and ‘0’ is not 
  the name of a set
- $0 \in 0$ – this is also ill-formed: ‘∈’ must be flanked by the name of a set on 
  the right and, again, ‘0’ is not such an expression

**Power sets**

Any set $S$ has subsets and we can collect these subsets together into a further set, 
which we call the *power set* of $S$, $\wp(S)$:

**Power set** The set consisting of all of the subsets of a set $A$ is called the *power 
set* of $A$, $\wp(A)$: $\wp(A) = \{x : x \subseteq A\}$.

Note that the empty set is a subset of every set. Consider some arbitrary set $S$. 
$\emptyset \subseteq S$, since $\forall x (x \in \emptyset \rightarrow x \in S)$. The antecedent of this conditional is always false, 
since $\emptyset$ has no members, so the universal claim is vacuously true.

For example, consider the set $\{a, b, c\}$. Its subsets are:
$\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}$.

The set of all of these subsets, the power set of $\{a, b, c\}$, is therefore:

$$\wp(\{a, b, c\}) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}.$$
There are several points to note here. First, every set is a subset of itself (though obviously not a proper subset!). Second, the empty set is a subset of every set. Third, the original set has 3 members and its power set has 8 members, i.e. $2^3$ members. Indeed, it is generally true that, if a set has $n$ members, its power set has $2^n$ members. When asked for the power set of an $n$-membered set, check that you follow this rule. If you have other than $2^n$ members, something has gone wrong! Check you’ve included the set itself and the empty set.

Order

We’ve said that order doesn’t matter to sets. Consider two-membered sets. We could call these unordered pairs, since the order of the members doesn’t matter: \{a, b\} = \{b, a\}. Sometimes we do care about order, though. For example, if I tell you that Jane and Ian finished first and second in a race, respectively, I don’t simply mean that the members of \{Jane, Ian\} finished first and second: I mean that Jane won and Ian was second.

To respect this, we introduce the idea of an ordered pair. The ordered pair \langle a, b \rangle \neq \langle b, a \rangle. For ordered pairs, \langle a, b \rangle = \langle c, d \rangle if, and only if, $a = c$ and $b = d$. We can think of ordered pairs as sets:

**Ordered pair** $\langle a, b \rangle = \{\{a\}, \{a, b\}\}$

Now that we have the definition of an ordered pair, we can of course go further. If we want to order three objects, we can use an ordered triple, \langle x, y, z \rangle. For four objects, we can use an ordered quadruple, \langle w, x, y, z \rangle. Generally, if we want to order $n$-many objects, we can use an ordered $n$-tuple, \langle $x_1...x_n$ \rangle.

If $A$ and $B$ are sets, then the set $A \times B$ is the Cartesian product of $A$ and $B$, whose elements are the ordered pairs such that the first member belongs to $A$ and the second member to $B$:

**Cartesian product** The Cartesian product of two sets $A$ and $B$, $A \times B = \{\langle x, y \rangle : x \in A \text{ and } y \in B\}$.

Let $A = \{0, 1\}$ and $B = \{1, c, d\}$. Their Cartesian product is:

$A \times B = \{\langle 0, 1 \rangle, \langle 0, c \rangle, \langle 0, d \rangle, \langle 1, 1 \rangle, \langle 1, c \rangle, \langle 1, d \rangle\}$

We can also talk about the Cartesian product of a set and itself. In this case, we say that the Cartesian product of $A$ and itself is $A^2$.

Complement

Our final set-theoretic operation is complement. Just as there is a clear analogy between union and disjunction, and intersection and conjunction, complement
corresponds to negation. For any sets $A$ and $B$, there is a set of objects that are members of $A$ but not $B$:

**Complement** $A - B = \{x : x \in A \land \neg x \in B\}$

(For non-membership, we can write either $\neg a \in S$ or $a \notin S$.)

We can now use our set-theoretic notation to write down expressions for various sets. Let $C$ and $N$ be the sets of all students at Churchill and Newnham, respectively. Also let $P$ be the set of philosophers and let $f$ be Frege. Then here are some other sets:

- $P \cap C$ – the set of Churchill philosophers
- $P \cap (C \cup N)$ – the set of all philosophers at either Churchill or Newnham
- $P \times C$ – the set of philosopher-Churchill student pairs
- $(P - \{f\}) \cup N$ – the set of all philosophers, except Frege, and Newnham students.

**Comprehension**

We have seen that Extensionality allows us to write $\{x : \phi(x)\}$ for the set of $\phi$s. What sort of property can $\phi$ be? The view that $\phi$ can be any property whatsoever is captured by the Principle of Comprehension:

**Comprehension** For any property $\phi$, the set $\{x : \phi(x)\}$ exists.

Comprehension tells us that there is the set of dogs (let $\phi$ be dog). It also tells us that there is an empty set (let $\phi$ be non-self-identity. And it tells us that there is a universal set (let $\phi$ be identity).

$\phi$ is also allowed to be a property of sets, such as having $\geq 4$ or more members. The sorts of things with members are sets and this instance of Comprehension tells us that there is a set of sets with 4 or more members. This set will include the set of members of the Beatles, the set of natural numbers, the set of pages in *Principia Mathematica*, the set of Oscar-winning films, and so on.

The Principle of Comprehension is false. We cannot comprehend just any property. In other words, some properties do not define sets. Why not? If they did, we would be able to prove a contradiction.

Consider the property being self-membered. By Comprehension, this defines the set of all sets that are members of themselves. What sort of set could be a member of itself? Again, consider the set of sets with 4 or more members. We specified 4
members of that set, so that set itself has 4 or more members. Hence, the set of sets with 4 or more members is a member of itself.

Of course, sets can also fail to be members of themselves. The set of dogs contains only dogs and, since no set is a dog, that set is not a member of itself.

**Russell’s Paradox**

Now consider the property being non-self-membered. By Comprehension, there is a set, the Russell set, \( \{x : x \notin x\} \). Unfortunately, there can be no such set. Let the Russell set be \( R \). Is \( R \) a member of itself? Well, if it is then it isn’t, since that is the condition for being a member of \( R \). So is it not a member of itself? Well, if it’s not, then it is, since again that’s the condition. So it’s a member of itself if, and only if, it isn’t, and that’s a contradiction.

Let’s go through that reasoning a little more slowly:

1. \( R \notin R \) – assumption
2. \( R \in R \) – 1, def. of \( R \)
3. \( R \in R \land R \notin R \) – 1,2
4. \( R \in R \) – 1-3, discharging 1
5. \( R \notin R \) – assumption
6. \( R \notin R \) – 5, def. of \( R \)
7. \( R \in R \land R \notin R \) – 5,6
8. \( R \notin R \) – 5-7, discharging 5
9. \( R \in R \land R \notin R \) – 4,8

This result is always known as Russell’s Paradox. All of the reasoning here is classically valid, so the problem must come from \( R \). \( R \) came from Comprehension, so we must reject (or at least restrict) Comprehension. The set theory characterised by Comprehension is known as naïve set theory. The version of Comprehension that lands it in trouble is called naïve (or unrestricted) Comprehension.

Here’s an analogy:

There is a barber who shaves all and only those people who do not shave themselves. Does the barber shave themself?
Let’s say that the barber does shave themself. Well, they only shave people who don’t shave themselves, so the barber doesn’t shave themself. So does the barber not shave themself? Well, they only shave people who don’t shave themself, so they do shave themself. And that’s a contradiction.

Now, it’s easy to say what’s gone wrong in the barber case. The story about the barber was incoherent so there can be no such barber. In the case of Russell’s Paradox, the answer is not so easy: it followed from our (naïve) theory of set. And the naïve theory really is very intuitive and natural. it is striking that such a conception is inconsistent.

**The iterative conception**

How should we respond to Russell’s Paradox? The usual answer is the iterative conception. You don’t need to know about this conception until Part II Mathematical Logic, so this brief sketch is just for enthusiasts. The iterative conception thinks of the set-theoretic universe as forming a V-shaped hierarchy. There are two guiding thoughts:

- Sets are formed in stages.
- At each stage, we form all possible collections of sets available from earlier stages.

Because there is an earliest stage at which each set is formed, and because a set’s members have already been formed, no set can be a member of itself. This is essentially how Russell’s Paradox is avoided.

## 2 Relations

**Intensions and extensions**

An important application of set theory is in treating properties and relations. We’ve been discussing properties casually so far, e.g. Comprehension tells us that there is a set of objects with the property *being a dog*. Now we should get clearer. You are familiar with the concept of a *predicate*. e.g:

```plaintext
_____ is a dog
_____ is a philosopher
```

Predicates express properties. Properties have intensions and extensions. Coextensive predicates can have different intensions. For example, the property *having three angles* and *having three sides* are coextensive (indeed, necessarily so):
they apply to precisely the same objects. But their intensions (or meanings) are clearly different: one is about angles, the other sides.

Sets provide a handy way of discussing the extensions of properties. The set \( \{ x : x \text{ is a dog} \} \) is the extension of the property being a dog. Two properties are coextensive just when their extensions are identical. The extension of having three angles and of having three sides are identical: \( \{ x : x \text{ has three angles} \} = \{ x : x \text{ has three sides} \} \).

Recall that predicates can have more than one gap, e.g.

\[
\begin{align*}
\_1 & \text{ loves } \_2 \\
\_1 & \text{ is taller than } \_2
\end{align*}
\]

These express relations: loving and being taller than, respectively. What are the extensions of relations? They can’t be sets of objects because relations don’t apply to objects individually but to pairs of objects. So can we use sets of pairs of objects? No, because that loses all sense of order. We want the relation shorter than to be distinct from the relation taller than, but the same pairs of objects fall under both. For example, Kylie Minogue is shorter than Nick Cave so the set \{Minogue, Cave\} would fall under the first. But the set \{Cave, Minogue\} would fall under the second. The problem, of course, is that \{Minogue, Cave\} = \{Cave, Minogue\}.

The solution is to use ordered pairs. After all, \( \langle \text{Minogue, Cave} \rangle \neq \langle \text{Cave, Minogue} \rangle \). The extensions of relations, then, are sets of ordered pairs. In particular, the extension of shorter than is \( \{ \langle x, y \rangle : x \text{ is shorter than } y \} \). Just as properties like unicorn can have empty extensions, so can relations, like shorter and taller than.

**Properties of relations**

Recall that when offering interpretations in FOL, you had to specify a domain. The same is true here: when we assign an extension to a relation, we specify a domain from which the objects are drawn. The reason for this will become clear when we consider some examples.

We are now in a position to discuss some of the properties of relations:

**Reflexivity** A relation \( R \) is reflexive on a domain just if everything in the domain bears \( R \) to itself, i.e. \( \forall x R xx \).

To judge whether a relation is reflexive on a given domain, you must ask yourself whether the result of assigning any object in that domain to both gaps in the relation results in a true claim. If it is true for every assignment, then that relation is reflexive on that domain. If there is at least one assignment making the claim false, it isn’t reflexive.
For example, consider the relation *x is the same height as y* on the domain of living people. Assigning Donald Glover to both *x* and *y* results in a true claim, since he is the same height as himself. Assigning Ariana Grande to both *x* and *y* also results in a true claim, since she is also the same height as herself. And soon you realise that *everyone* is the same height as themself, hence the relation is indeed reflexive on that domain.

Consider *x is strictly taller than y* on the same domain. Take Donald Glover again: assign him to both *x* and *y*. The result is false, since he isn’t strictly taller than himself. And that’s enough to show that the relation is not reflexive.

For example, *x is no bigger than y* is a reflexive relation on any domain. If a relation is reflexive on any domain, we can say it is reflexive *simpliciter*. The relation *x and y are both right-handed* will be reflexive on a domain containing only right-handed people, but it is of course not reflexive *simpliciter*.

**Anti-reflexivity** A relation *R* is anti-reflexive on a domain just if *nothing* in the domain bears *R* to itself, i.e. $\forall x \neg Rxx$.

For example, the relation *x is at least one foot taller than y* is anti-reflexive on any domain. Some relations are neither reflexive nor anti-reflexive on a given domain. The right-handed example works again: on a domain containing both left- and right-handed people, *x and y are both right-handed* will have false instances (the left-handed people), hence it’s not reflexive on that domain, and true instances (the right-handed people), hence it’s not anti-reflexive on that domain.

Note that Anti-reflexivity is not the negation of Reflexivity, $\neg \forall xRx$. When this condition obtains, we will call the relation *irreflexive*. Can a relation be *both* reflexive and anti-reflexive? Only if the domain is empty: then they will both be vacuously true in the way that ‘everything is a unicorn’ and ‘nothing is a unicorn’ both are. But recall that we insisted in forallx that domains be non-empty. We will continue with that insistence here.

The next property of relations we will consider is **symmetry**:

**Symmetry** A relation *R* is symmetric on a domain just if, whenever *x* bears *R* to *y*, *y* also bears *R* to *x*, i.e. $\forall x \forall y (Rxy \rightarrow Ryx)$.

The relation *is married to* is symmetric on the domain of people since, if someone *x* is married to someone *y*, it follows that *y* is also married to *x*. Note that, since we can quantify over *any* *x* and *y* in the definition of Symmetry, we can derive the converse conditional, hence we can derive the biconditional.

As with Reflexivity, we can consider relations of *non*-symmetry. In the case of symmetry, there are two claims we could intend here:
Asymmetry A relation $R$ is asymmetric on a domain just if, whenever $x$ bears $R$ to $y$, $y$ does not bear $R$ to $x$, i.e. $\forall x \forall y (Rxy \rightarrow \neg Ryx)$.

Anti-symmetry A relation $R$ is antisymmetric on a domain just if, whenever $x$ bears $R$ to $y$ and $y$ bears $R$ to $x$, $x$ and $y$ are identical, i.e. $\forall x \forall y ((Rxy \land Ryx) \rightarrow x = y)$.

On the domain of people, the relation $x$ is the father of $y$ is asymmetric, since no one is their own father. Of course, if a relation is asymmetric, then we will never have both $Rxy$ and $Ryx$, so the relation will also be vacuously anti-symmetric. The converse is not true: there are anti-symmetric relations that are not also asymmetric. An example is $x$ is greater than or equal to $y$ on the domain of natural numbers. If two numbers are either greater than or equal to each other, then they must be the same number. But the same relation is not asymmetric. For example, 2 is greater than or equal to 2, but it obviously doesn’t follow that 2 is not greater than or equal to 2. For this reason, an asymmetric relation cannot be reflexive.

These considerations bring out another important point: relations can be empty. Because Symmetry, Asymmetry and Anti-symmetry are all conditional, an empty relation will have all three properties vacuously. Not, however, that an empty relation isn’t vacuously reflexive. Reflexivity demands that every object-object pair $\langle x, x \rangle$ in the domain is in the extension of the given relation. But nothing is in the extension of the empty relation, and we know that there is something in the domain, so empty relations are not reflexive. Anti-reflexivity, on the other hand, can hold vacuously.

Now consider:

Transitivity A relation $R$ is transitive on a domain just if, whenever $x$ bears $R$ to $y$ and $y$ bears $R$ to $z$, $x$ bears $R$ to $z$, i.e. $\forall x \forall y \forall z ((Rxy \land Ryz) \rightarrow Rxz)$.

Our example of an antisymmetric relation, $x$ is the father of $y$, also fails to be transitive on the domain of all people ever. The relation, $x$ is an ancestor of $y$, on the other hand, is transitive on the same domain.

The intuitive idea behind transitivity is of a relation holding down chains. Remember Derek Parfit’s distinction between psychological continuity and connectedness in ‘Personal Identity’ (1971). Let the domain again be all people ever. Psychological connectedness is the holding of a range of psychological states such as memories in common. It is not transitive, since this close psychological connection fails to hold down chains: I will have quite different psychological states to myself as a child. Psychological continuity, however, is transitive since it is the overlapping chain of psychological connectedness. I am psychologically continuous with myself as a child: hence, in Parfit’s language, that child’s surviving as me.

Again, we find an inverse to transitivity:
**Anti-transitivity** A relation $R$ is *anti-transitive* on a domain just if, whenever $x$ bears $R$ to $y$ and $y$ bears $R$ to $z$, $x$ does not bear $R$ to $z$, i.e.
\[ \forall x \forall y \forall z ((Rx y \land Ryz) \rightarrow \neg Rxz). \]

As with anti-reflexivity, this is not the *negation* of transitivity, which we will call *intransitivity*. A relation can be both transitive and anti-transitive. The empty relation is another example of this, since these definitions are again conditionals which can hold vacuously.

**Equivalence Relations**

We have met the following properties of relations: Reflexivity, Symmetry and Transitivity. When a relation has all three of these properties we call it an *equivalence relation*.

Consider the relation *$x$ is exactly as tall as $y$* on the domain of living people. It is clearly *reflexive*, since everyone is exactly as tall as themself. It is also *symmetric*: if $x$ is exactly as tall as $y$, then $y$ is exactly as tall as $x$. And it is *transitive*: if $x$ is exactly as tall as $y$, and $y$ as $z$, then $x$ is exactly as tall as $z$. Hence *$x$ is exactly as tall as $y$* is an equivalence relation. Other examples on the same domain are:

- $x$ was born in the same town as $y$
- $x$ is exactly as old as $y$
- $x = y$

In contrast, *$x$ is married to $y$* on the same domain is not an equivalence relation. This is because it fails to be reflexive, since no one is married to themself. The relation *$x$ is a brother of $y$* on the same domain is also not an equivalence relation, since it fails to be symmetric: $x$ may be a brother of $y$ and $y$ be a *sister* of $x$.

Finally, *$x$ was born or died in the same town as $y$* on the domain of all people ever is not an equivalence relation because it fails to be transitive: $x$ and $y$ may have only their place of birth in common, $y$ and $z$ only their place of death in common, and $x$ and $z$ neither place in common.

**Partitions**

Equivalence relations have the interesting property of *partitioning* a domain. A partition is a way of dividing up the members of a domain into sectors such that every member belongs to exactly one sector. The division is *exclusive* (nothing belongs to more than one sector) and *exhaustive* (every member belongs to at least one sector).

As with many of the notions under discussion, set theory provides a helpful way of discussing partitions. Let $S$ be a set and $P$ be a set of subsets of $S$, i.e.
\[ \forall x (x \in P \rightarrow x \in \phi(S)). \]
**Partition**  $P$ is a partition on $S$ just if:

$$\forall x(x \in S \rightarrow \exists y(y \in P \land x \in y \land \forall z((x \in z \land z \in P) \rightarrow y = z)))$$

The sectors into which a partition carves a domain are equivalence classes. We will call these the equivalence classes of the relation that does the partitioning.

Consider the relation $x$ was born in the same country as $y$ on the domain of living people. This is an equivalence relation because it is reflexive, symmetric, and transitive. It therefore partitions the domain into equivalence classes. In this case, one class will include all of the people born in the UK, one all the people born in France, and so on for every country.

We have said that $x$ is exactly as tall as $y$ on the domain of living people is an equivalence relation. This relation partitions the domain into equivalence classes of people of the same height. There is a class of people who are exactly 5 feet tall, one of people 6 feet tall, and so on.

For their logicist projects, Russell and Frege used similar techniques to define the notion of cardinal number (intuitively, numbers that measure the sizes of sets). The equivalence relation they used as $x$ and $y$ are equinumerous on the domain of sets. This partitions the sets into those with no members, those with one member, and so on.

Finally, consider the identity relation. We have said that this is an equivalence relation on the domain of everything so it must partition the members of that domain into equivalence classes. What are these classes? Each contains only one member, since nothing is identical to anything but itself.

**Converse**

If $R$ is a relation, then we will say that the converse of $R$, $R'$, is the relation that holds between $x$ and $y$ just if $R$ holds between $y$ and $x$:

**Converse** Relation $R'$ is the converse of relation $R$ just if $\forall x \forall y (Rxy \leftrightarrow R'yx)$.

If $R$ is the relation $x$ is bigger than $y$, then $R'$ is the relation $x$ is smaller than $y$. Intuitively, the converse is the opposite of the relation in question.

Some relations are their own converses, e.g. $x$ is a sibling of $y$.

**Ancestral**

If $R$ is a relation, we will say that the ancestral of $R$, $R^*$, is the relation that holds between $x$ and $y$ just if for some $z, w, v, u, ...$, we have $Rxz \land Rzw \land ... \land Rvu \land Ruy$. Intuitively, there is a chain from $x$ to $y$ such that each link in the chain bears that relation to the next. If $R$ is $x$ is a parent of $y$, then $R^*$ is $x$ is an ancestor of $y$.  

14
To link this once again to personal identity, we know that the memory relation \( x \) remembers being \( y \) on the domain of living people is not transitive, since I may not remember being a time slice of myself as a child. But identity is transitive on that domain, so \( x \) and \( y \) are the same person cannot be the memory relation. But take the ancestral of the memory relation and the claim is more plausible.

**Complexity**

So far, the relations we have considered have all been intuitive and natural-looking. This needn’t be the case, however. Recall that a *predicate* is defined very broadly: *any* deletion of names from a sentence and replacement with gaps is a predicate. Hence predicates can look very unnatural, e.g. ‘\( x \) is the lecturer for Formal Logic in 2018/19 and \( y \) is the lecturer for Sets, Relations and Probability in 2018/19’. That expresses a relation that holds only between the pair of Tim and me. It’s hard to imagine such a relation being of any real interest, but it’s a relation nevertheless.

The generous definition also entails that predicates can have logical complexity, e.g. ‘\( x \) is Donald Glover \( \iff \) \( y \) is Ariana Grande’. That’s a perfectly reasonable predicate, expressing a perfectly reasonable relation, even though it contains logical complexity.

When logical complexity is involved, it can be difficult (and certainly unintuitive) to judge its properties on a given domain. Let’s work through this example on the domain of currently living people. My first tip is not to attempt to have intuitions about the case (at least not until you’re very comfortable with these) but instead to write down the long sentence that must be true if the relation is to have the property in question.

Let’s start with reflexivity. The following sentence has to be true on the domain of currently living people if the relation is reflexive:

\[
\forall x (x \text{ is Donald Glover } \iff x \text{ is Ariana Grande})
\]

Is that sentence true? Well, it certainly holds of all of us, since none of us are Donald Glover and none of us are Ariana Grande. How about Donald Glover himself? It is false of him, since he is Donald Glover but he is not Ariana Grande. By analogous reasoning, it is false of Ariana Grande. So the universal has false instances. So the sentence is false. The relation is not reflexive.

What about symmetry? To be symmetric on the same domain, this sentence must be true:

\[
1 \forall x \forall y ((x \text{ is Donald Glover } \iff y \text{ is Ariana Grande}) \rightarrow (y \text{ is Donald Glover } \iff x \text{ is Ariana Grande}))
\]

Is this true? It’s a conditional so it’s only false when the antecedent is true and the consequent false. Hence it is only false when, for some value of \( x \) and \( y \):
2 \( x \) is Donald Glover \( \leftrightarrow \) \( y \) is Ariana Grande
3 \( \neg(y \) is Donald Glover \( \leftrightarrow x \) is Ariana Grande) 

There are only two ways that 2. can be true:

4 \( x = \)Donald Glover and \( y = \)Ariana Grande
5 \( x \neq \) Donald Glover and \( y \neq \) Ariana Grande 

Let’s start with 4. If that’s the case, then 3 is false. Why? Because both sides of the biconditional are false, so the biconditional itself is true, so the whole sentence is false. So 4 is not a counterexample to 1.

We are left with 5. We need \( x \) to be anyone by Donald Glover and \( y \) anyone but Ariana Grande. Let’s say they are Tim (not Donald Glover) and me (not Ariana Grande) But then both sides of the biconditional in 3 are false again, so the biconditional is true again, and its negation false. No counterexample.

But we haven’t exhausted the possibilities in 5, since there are many ways of failing to be Donald Glover and Ariana Grande. In particular, let \( x = \)Ariana Grande and \( y = \)Tim. Now the biconditional in 3 is false, since its LHS is false (Tim \( \neq \) Donald Glover) but its RHS is true (Ariana Grande = Ariana Grande). So it’s negation is true. This is a valuation that makes 2 (the antecedent of the conditional in 1) true but 3 (the consequent of the conditional in 1) false. So 1 has a false instance, so 1 is false. The relation is not symmetric.

That’s a slow working-through of how we would decide whether a complex relation like this has some property or other. It’s difficult to have intuitions about so you should work through something like this process. I leave it as an exercise to judge whether the same relation possesses the other properties discussed, on the same domain.

**Examples**

Questions on relations typically take one of two forms:

- Here are some relations and a domain. Determine whether they have these properties on that domain (and there may be some unfamiliar property thrown in).

- Here are some properties (again, there may be an unfamiliar property). Provide relations (and domains, if not specified) that exhibit those properties.

Let’s take the two in turn:
1 On the domain of all people ever, determine whether the relation \( x \) is Frege or \( y \) is Russell is reflexive, whether it is symmetric and whether it is transitive.

2 Let’s say that a relation is Euclidean on a domain just if
\[
\forall x \forall y \forall z ((Rxy \land Rxz) \rightarrow Ryz).
\]
Determine whether, on the domain of living people, the relation \( x \) loves \( y \) only if \( y \) loves \( x \) is reflexive, whether it is symmetric, whether it is transitive and whether it is Euclidean.

Of the second sort, you might get questions like:

3 Give an example of a relation that is reflexive and transitive but not symmetric on a domain.

4 Give an example of a relation that is symmetric and transitive, but not reflexive on a domain.

5 Give an example of a relation that is neither reflexive, symmetric, transitive, not Euclidean on a domain.

3 Probability

Why study probability?

Probabilistic reasoning is rife in philosophy. In the philosophy of science, we discuss the extent to which theories are confirmed probabilistically. In metaphysics, probabilistic accounts of causation are popular. In the philosophy of language, there are probabilistic accounts of conditionals. Probability is central to quantum mechanics, the philosophy of biology, philosophy of statistics and philosophy of social science. Probability is also a central component of decision theory, game theory and their applications to e.g. moral and political philosophy. In the philosophy of mind, various mental states are often represented as probabilities (degrees of belief). And, of course, there is the philosophy of probability itself, which we’ll touch on later. In this part of the course, we will focus on the probability calculus but also have something to say of its philosophical interpretation.

Outcomes

We will be assigning probabilities to sets of outcomes. There are other candidates in the literature, such as propositions and sentences, but sets of outcomes seem the most natural for various reasons.

Let’s say that the probability of the Conservatives winning the next election is 40%. It seems natural to regard a Conservative victory as a set of outcomes since it includes the case where they win by 1 seat, by 100 seats, etc, and these are
clearly different states of affairs. The fineness of grain here will normally be settled by context.

The set of possible outcomes in which we are interested will be called the outcome space, \( V \). In other places, you will find this referred to as the reference set or the sample space.

For example, if we throw a fair 6-sided die, there are 6 possible outcomes:
\[ V = \{1, 2, 3, 4, 5, 6\} \]. There are 6 possible outcomes but there are more than 6 sets of outcomes, which we will call events.

For example, we may want to know the probability of rolling an even number, or of rolling either a 1 or a 5. The first event would be represented as \( \{2, 4, 6\} \) and the second as \( \{1, 5\} \). How many events are there? For outcome space \( V \), the events will be the members of \( \phi(V) \). So, for an outcome space with \( n \) members, we know that there will be \( 2^n \) events. The set of events, \( \phi(V) \) we will call the field, \( F \), of \( V \).

The field contains the certain event, \( \{1, 2, 3, 4, 5, 6\} \) and the impossible event, \( \emptyset \).

Also, note that we can use our other set-theoretic operations: if \( X \in F \) and \( Y \in F \), then \( X \cap Y \in F \) and \( X \cup Y \in F \). We’ll also write \( X^c \) for \( V - X \), i.e. the set of outcomes other than those that are members of \( X \).

Kolmogorov axioms

We can now define a probability function as a function, \( Pr \), that assigns numbers to members of \( F \) and obeys the following Kolmogorov axioms for any \( X, Y \in F \):

1. \( Pr(V) = 1 \)
2. \( Pr(X) \geq 0 \)
3. If \( X \cap Y = \emptyset \), then \( Pr(X \cup Y) = Pr(X) + Pr(Y) \)

With just these in place, we are already in a position to prove an easy theorem: \( Pr(X) + Pr(X^c) = 1 \).

By definition, \( X^c = V - X \).

So \( X \cup X^c = V \).

By axiom 1, \( Pr(V) = 1 \).

Substituting, \( Pr(X \cup X^c) = 1 \).

By axiom 3, since \( X \cap X^c = \emptyset \), \( Pr(X \cup X^c) = Pr(X) + Pr(X^c) \).

Substituting, \( Pr(X) + Pr(X^c) = 1 \).

\(^1\)Some texts will write \( X^* \) for \( V - X \). I am not using this, to avoid confusion with ancestral.
Principle of indifference

Now we can assign probabilities to events. First, we assign probabilities to outcomes. The probability of rolling a 6 on a fair 6-sided die is, of course, $\frac{1}{6}$. Why? We might be tempted to say something about the physical symmetry of the die: it is a fair, evenly weighted die and so of course it will come up 6 on 1 in 6 rolls.

This is not the reason that we will give. Rather, the assignment of probabilities to outcomes is due to their frequencies. The probability of rolling a 6 is $\frac{1}{6}$ because it has in fact come up 6 about one-sixth of the time.

This point is important in light of Bertrand’s paradoxes. We may think that we can assign probabilities a priori according to the principle of indifference: whenever there is no evidence favouring one possibility over another, each should be assigned the same probability as the others. This is at the heart of the classical interpretation of probability.

Unfortunately, the principle is inconsistent. Suppose I tell you only that a car travelled 100 miles at an average speed of between 50 and 100mph. What is the probability that the average speed was between 75 and 100mph? You may be tempted to say 0.5, because the ranges 50-75mph and 75-100mph are equally possible.

But now I reformulate the problem: a car took between 1 and 2 hours to travel 100 miles. This is intuitively the same situation. What is the probability that the journey took between 1 hour and 1\frac{1}{3} hours? You may be tempted to say $\frac{1}{3}$ since the ranges $1 - 1\frac{1}{3}, 1\frac{1}{3} - 1\frac{2}{3}, 1\frac{2}{3} - 2$ hours are equally likely.

This is a problem: I have asked equivalent questions and the principle of indifference has delivered incompatible answers. This interpretation seems to make probability far too sensitive to description.

Frequentism

Instead, we will follow the frequentist interpretation. On this view, our observations of the two sorts of case considered here are the very same, and so they receive the same answer. Of course, there are problems. Which observations do we count? How many observations do there have to be? What are events? If events are unique, how can they be repeated?

This story is also open to the usual Humean worries about induction. We know that past experience does not make the future certain so it also doesn’t make it likely. The procedure is, then, for the usual Humean reasons, irrational.

To return to the die example, we will say that $Pr(\{6\}) = \frac{1}{6}$. Here, we assign a probability to a singleton since we want it to be a set. Similarly, for a standard deck of cards (52 different cards; 13 in each of 4 suits), the probability of drawing the king of clubs is $\frac{1}{52}$. In other words, $Pr(\{KC\}) = \frac{1}{52}$. 

19
We can also work out the probability of other events. What is the probability that a card drawn from a standard deck of cards is either the ace of hearts or an even spade? We need to consider the relevant subset of $V$. Let’s call it $T = \{2S, 4S, 6S, 8S, 10S, AH\}$. We want to know $Pr(T)$. From the third Kolmogorov axiom,

$$Pr(T) = Pr(\{2S\}) + Pr(\{4S\}) + Pr(\{6S\}) + Pr(\{8S\}) + Pr(\{10S\}) + Pr(\{AH\}) = \frac{6}{52} = \frac{3}{26}$$

**Conditional probability**

We are now in a position to define *conditional probability*:

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$$

We pronounce ‘$Pr(A|B)$’ as ‘the probability of $A$ given $B$’. It is of course undefined when $Pr(B) = 0$. The most natural interpretation of this definition is statistical: if $Pr(A)$ measures the proportion of $A$s within a population $V$, then $Pr(A|B)$ measures the proportion of $A$s amongst the $B$s in the same population.

For example, what is the probability that a fair 6-sided dice comes up 2, given that it comes up even? First, $V = \{1, 2, 3, 4, 5, 6\}$. The outcomes where the die is even are $E = \{2, 4, 6\}$. The outcome that the die is 2 is $T = \{2\}$. So, applying the formula:

$$Pr(T|E) = \frac{Pr(T \cap E)}{Pr(E)}$$

But note that $T \cap E = T$ because $T \subseteq E$. Now $Pr(T \cap E) = Pr(T) = \frac{1}{6}$. And $Pr(E) = \frac{1}{2}$. So:

$$Pr(T|E) = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

**Independence**

If $X, Y \in F$, then $X$ and $Y$ are *probabilistically independent* for a given probability function, $Pr$, if $Pr(X|Y) = Pr(X)$. Otherwise, they are *probabilistically dependent*. These relations are both symmetric.

For example, rolling a 1 and rolling a 2 on a fair die are independent events. But rolling a 2 and rolling an even number are dependent events. In the example from the previous section, we saw that $Pr(T|E) = \frac{1}{3}$, whereas $Pr(T) = \frac{1}{6}$. So $Pr(T|E) \neq Pr(T)$, so $T$ and $E$ are dependent.

This all makes intuitive sense. The difference corresponds to the intuitive distinction between events that do, and that do not, ‘make a difference’ to another event. Statistical probabilistic dependence can hold between events that stand in no causal connection.
We can rearrange the formula for conditional probability to get:

\[ Pr(A \cap B) = Pr(A|B) \times Pr(B) \]

This gives us a means of calculating \( Pr(A \cap B) \) (intuitively, \( A \) and \( B \)) when we have \( Pr(A|B) \) and \( Pr(B) \).

Remember that \( A \) and \( B \) are **probabilistically independent** when \( Pr(A|B) = Pr(A) \), and that this relation is symmetric. Hence, when \( A \) and \( B \) are probabilistically independent, \( Pr(A \cap B) = Pr(A) \times Pr(B) \). This is not true when they are **dependent**.

For example, suppose we know that exactly 50% of the population are female and that 30% of females have blue eyes. What is the probability that a randomly selected person is female and blue-eyed?

Let \( F \) be the event of selecting a female and \( B \) the event of selecting someone with blue eyes. We are given that \( Pr(F) = 0.5 \) and \( Pr(B|F) = 0.3 \). Rearranging the conditional probability formula and substituting:

\[ Pr(B \cap F) = Pr(B|F) \times Pr(F) = 0.3 \times 0.5 = 0.15 \]

**Repeated trials**

So far, we have only considered single die rolls or card draws, but we may well be interested in repetition of these. Suppose that the outcome space for some trial is \( V \). If the possible outcomes are the same in both cases, we can consider the repeated trial as a single trial and use ordered pairs to capture the outcomes. If \( V \) is the outcome space for each trial, it is natural to treat the outcome space of the repeated trial as \( V \times V = V^2 = \{ \langle x, y \rangle : x \in V \land y \in V \} \). Order is of course crucial here, so that we can distinguish events that are intuitively distinct.

The approach we take to such cases will depend on whether the trials are **probabilistically independent** or not. The former is somewhat simpler, so let’s start there.

**Independent trials**

Suppose you toss a coin twice and you want to know the probability of tossing at least one heads. These are clearly independent events, since each coin toss has no bearing on the other.

The first step with questions of this sort is to work out the outcome space. The outcome space of each toss considered individually is \( V = \{ H, T \} \). The outcome for both considered jointly is therefore \( V^2 = \{ \langle H, H \rangle, \langle H, T \rangle, \langle T, H \rangle, \langle T, T \rangle \} \). So there are four possible outcomes.

The next step is to consider the relevant formula. To do this, it is helpful to consider the set whose probability we are after. Here, it is
\{\langle H, H \rangle \} \cup \{\langle H, T \rangle \} \cup \{\langle T, H \rangle \}. What formula allows us to calculate the probability of such sets? The third Kolmogorov axiom:

For $X, Y \in F$, if $X \cap Y = \emptyset$, then $Pr(X \cup Y) = Pr(X) + Pr(Y)^2$

This applies because the three sets in question are clearly disjoint: they each have only one member and it is unique in each case. So we need to calculate:

$$Pr(\{\langle H, H \rangle \}) + Pr(\{\langle H, T \rangle \}) + Pr(\{\langle T, H \rangle \}) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$$

**Dependent trials**

An example of repeated trials where we have probabilistic dependence is the repeated card draws from a deck without replacement. The probability of the second draw will depend on the first, since a card has been removed from the deck. Intuitively, the first draw makes a difference to the second, since the deck is reduced. Nothing like this was the case with the coin tosses.

The outcome space, $V$, for a single card draw has 52 members. In the case of two card draws without replacement, the outcome space cannot be $V^2$, since we cannot remove the same card twice. Rather, it is $V^2 - \{\langle x, x \rangle : x \in V \}$. How many members does this set have? $52 \times 51$, each of whose members has equal probability.

Suppose we want to know the probability that at least one of our two card draws (without replacement) is an ace. It will help to first label some events to stop the notation getting too cumbersome:

- **A** The first card is an ace and the second is not.
- **B** The second card is an ace and the first is not.
- **C** Both cards are aces.

We want $Pr(\{A \cup B \cup C\})$. First note that the intersection of any pair of these sets is empty: every member of $A$ has an ace in the first place, removing it from $B$, and a non-ace in second, removing it from $C$. Similar reasoning shows that all the sets are disjoint. So:

$$Pr(\{A \cup B \cup C\}) = Pr(A) + Pr(B) + Pr(C)$$

What are these probabilities?

- **A** has $4 \times 48$ members. So $Pr(A) = \frac{4 \times 48}{52 \times 51}$

---

\(^2\)When the question is not about conditional probability, this will usually be the required formula.
By parity of reasoning, $Pr(B) = \frac{4 \times 48}{52 \times 51}$

The set $C$ has $4 \times 3$ members. So $Pr(C) = \frac{4 \times 3}{52 \times 51}$. So our answer is:

$$Pr(A \cup B \cup C) = \frac{(4 \times 48) + (4 \times 48) + (4 \times 3)}{52 \times 51} = \frac{99}{13 \times 51} = \frac{33}{13 \times 17} = \frac{33}{221}$$

**Conditional probability of repeated trials**

So far, we’ve not considered how to combine repeated trials with conditional probability. Again, let’s first consider independent trials and then move on to the more complex dependent ones.

**Independent trials**

Let’s again take coin tosses as our example here. Let’s say you toss a fair coin three times and you want to know the probability that they are all heads given that at least one of them is. Here, taking $V$ as the outcome space for a coin toss, the total outcome space will be:

$$V^3 = \{\langle H, H, H \rangle, \langle H, H, T \rangle, \langle H, T, H \rangle, \langle H, T, T \rangle, \langle T, H, H \rangle, \langle T, H, T \rangle, \langle T, T, H \rangle, \langle T, T, T \rangle\}$$

These 8 outcomes all have equal probability: $\frac{1}{8}$.

We want to calculate:

$$Pr(\{\langle H, H, H \rangle\} | \{\langle H, H, H \rangle, \langle H, H, T \rangle, \langle H, T, H \rangle, \langle H, T, T \rangle, \langle T, H, H \rangle, \langle T, H, T \rangle, \langle T, T, H \rangle, \langle T, T, T \rangle\})$$

From the formula for conditional probability, this value is:

$$Pr(\{\langle H, H, H \rangle\} \cap \{\langle H, H, H \rangle, \langle H, H, T \rangle, \langle H, T, H \rangle, \langle H, T, T \rangle, \langle T, H, H \rangle, \langle T, H, T \rangle, \langle T, T, H \rangle, \langle T, T, T \rangle\})$$

Notice again that the top quantity reduces to $Pr(\{\langle H, H, H \rangle\}) = \frac{1}{8}$. And this needs to be divided by $\frac{7}{8}$:

$$\frac{1}{\frac{7}{8}} = \frac{1}{\frac{7}{8}} = \frac{8}{7}$$

Suppose we now want to calculate the probability that all three tosses are heads, given that the first is heads. In this case:

$$Pr(\{\langle H, H, H \rangle\} | \{\langle H, H, H \rangle, \langle H, H, T \rangle, \langle H, T, H \rangle, \langle H, T, T \rangle\}) = \frac{1}{\frac{8}{7}} = \frac{1}{4}$$

These results should give us pause. We have shown that the probability of tossing three heads given that at least one is heads is $\frac{1}{7}$. And the probability of tossing three heads given that the first one is heads is $\frac{1}{4}$. So the latter is almost twice as

---

<sup>3</sup>It is often useful to keep the numbers in these unexpanded forms at first, since they will often cancel.
likely as the first. But that seems counterintuitive: if we know that at least one coin toss is a heads, how could it matter that the *first* one is heads?

The lesson is that conditional probabilities can often yield surprising results. It is crucial that you apply the formula for conditional probability and don’t just offer ‘intuitive’ reasoning, which is normally a disaster.

Look at the calculations we made. In both cases, the numerator is the same. So our explanation for the difference must be in the denominator, which in the first case is $\frac{7}{8}$ and in the second $\frac{4}{8}$. In the second, our value is almost half. Why is this? Remember what this is measuring: in the first case, the number of outcomes with at least one head; in the second, the number of outcomes with a head in first place. There are fewer of these (4 rather than 7) so they are less likely to occur.

**Dependent trials**

Finally, let’s consider conditional probabilities and *dependent* trials. And again, let’s take card draws without replacement as our example. Let’s calculate the probability that if two cards are drawn from a pack without replacement, both are aces given that one is an ace. Again, let’s label some events:

- **$A$** Both cards are aces.
- **$B$** At least one of the cards is an ace.

From our earlier example\(^4\), we know that $Pr(A) = \frac{4 \times 3}{52 \times 51} = \frac{1}{221}$. Similarly, we know that $Pr(B) = \frac{33}{221}$. We also know that $A \cap B = A$ so:

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{Pr(A)}{Pr(B)} = \frac{1}{221} \cdot \frac{221}{33} = \frac{1}{33}$$

Now let’s calculate the probability that they are both aces given that one of them is the ace of spades. Let:

- **$C$** One of the cards is the ace of spades.
- **$C_1$** The first card is the ace of spades (and the second is not). 51 outcomes
- **$C_2$** The second is the ace of spades (and the first is not). 51 outcomes
- **$A \cap C_1$** The first is the ace of spades; the second is another ace. 3 outcomes
- **$A \cap C_2$** The second is the ace of spades; the first is another ace. 3 outcomes

\(^4\)It’s always worth looking at previous results when answering these questions. Often, earlier results will save time in answering later questions.
We want $Pr(A|C)$ so we need $Pr(A \cap C)$ and $Pr(C)$. These are:

$$Pr(C) = Pr(C_1) + Pr(C_2) = \frac{102}{52 \times 51}$$

$$Pr(A \cap C) = Pr(A \cap C_1) + Pr(A \cap C_2) = \frac{6}{52 \times 51}$$

$$Pr(A|C) = \frac{6}{102} = \frac{1}{17}$$

Again, we might be surprised by these results. The probability that both cards are aces given that one is the ace of spades is almost double the probability that both are aces given that at least one is an ace. Why should this extra information be relevant? Again, conditional probabilities can be surprising.

**Subjective probability**

We have been working on the assumption that probabilities measure frequencies. We might call such probabilities objective. Consider quantum mechanics: certain events at this level are absolutely unpredictable. For example, a radium atom may decay in a given interval or it may not. Decaying and non-decaying atoms cannot be distinguished: all we can say is that any given atom has a certain probability of decaying. A radium atom has a 50% probability of decaying over an interval of 1,602 years – the half-life of the atom. This probability seems, in some sense, to be ‘in the world’.

On the other hand, we might take probabilities as measuring our confidence in a particular outcome. Such probabilities would be subjective. The Bayesian claim is that, after you learn some new piece of evidence $E$, you new confidence in any hypothesis is your old $Pr(H|E)$. In other words, you should be constantly updating your confidence in hypothesis in light of new evidence, using conditional probability.

**Bayes’ Theorem**

Let $H$ be some hypotheses and $E$ be some piece of evidence. We know from the conditional probability formula that:

$$Pr(H|E) = \frac{Pr(H \cap E)}{Pr(E)}$$

$$Pr(E|H) = \frac{Pr(E \cap H)}{Pr(H)}$$

Noting that intersection is symmetric and combining, we get:

$$Pr(H|E) \times Pr(E) = Pr(E|H) \times Pr(H)$$

Rearranging, we get Bayes’ Theorem:

$$Pr(H|E) = \frac{Pr(E|H) \times Pr(H)}{Pr(E)}$$
In other words, if we know the probability of some hypothesis, and of some piece of
evidence, and of that evidence given that hypothesis, we can calculate the
probability of that hypothesis given that evidence. One upshot is that, if some
hypothesis predicts an observation that was very unlikely in advance, then actually
making that observation is very strong evidence for the hypothesis.

Suppose I hold some bizarre conspiracy theory $H$ according to which the end of
the world will occur this week. And suppose that, on the basis of $H$, I predict $E$,
that there will be widespread earthquakes all over the world tomorrow. Today, we
might think that $Pr(H)$ and $Pr(E)$ are very low. We might also think that
$Pr(E|H)$ is very high, for argument’s sake. This is because, although we don’t
think my bizarre theory has any plausibility, we’re happy to allow that if it’s
correct, then there will be widespread earthquakes. Given all this, if we observe
widespread earthquakes tomorrow, then the likelihood of the hypothesis rises
enormously (by a factor of $\frac{1}{Pr(E)}$).

The Monty Hall puzzle

The Monty Hall puzzle is a famous problem in probability based on the American
game show *Let’s Make a Deal*, hosted by Monty Hall. You’re on a game show and
can choose one of three doors. You know that behind one door is a car and behind
the other two are goats. You want the car. You choose, for argument’s sake, door
1. Monty Hall - the presenter, who knows what is behind each door - tells you that
the car is *not* behind door 3, and opens it to reveal a goat. If you happen to have
picked the correct door already, he’ll pick a remaining door at random. He asks
you if you want to switch to door 2 or stick with door 2. Should you switch?

Let $H$ be the event that you initially picked the correct door. Let $E$ be the event
that Monty Hall opened door 3. $Pr(H) = \frac{1}{3}$; $Pr(E) = \frac{1}{2}$ and $Pr(E|H) = \frac{1}{2}$. Now,
plugging all this into Bayes’ Theorem:

$$Pr(H|E) = \frac{Pr(E|H) \times Pr(H)}{Pr(E)} = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

So you should be twice as confident that the car is behind door 2. You should
switch! (Assuming that you don’t really want a goat.)

This is all highly counterintuitive. What’s going on? When you made your original
choice, your chance of being correct (the car is behind door 1) was clearly $\frac{1}{3}$. So
your chance of being incorrect (the car is behind door 2 or 3) was $\frac{2}{3}$. But, Monty
having eliminated door 3, the $\frac{2}{3}$ probability is now, if you like, all behind door 2.

Here’s another explanation. Monty is a TV pro with a nose for tension, so he
would never open the door you picked. So the probability that you were initially
correct hasn’t changed, it’s still $\frac{1}{3}$. Now, door 3 having being eliminated, there’s
only one live option, that the car is behind door 2. And that probability must be
$1 - \frac{1}{3} = \frac{2}{3}$. You should switch!

Bayes’ Theorem (second version)
Bayes’ Theorem (in the version so far considered) allows us to calculate $Pr(H|E)$ on the basis of knowledge of $Pr(E|H)$, $Pr(H)$ and $Pr(E)$. Sometimes, however, we may be unaware of $Pr(E)$. Perhaps we only know $Pr(E|H)$ and $Pr(E\overline{H})$. For example, imagine you have test whose reliability you know and you want to determine – for a given proposition – its likelihood given a positive result.

Clearly:
$$E = (E \cap H) \cup (E \cap \overline{H})$$

Equally clearly:
$$(E \cap H) \cap (E \cap \overline{H}) = \emptyset$$

Hence, by axiom 3:
$$Pr(E) = Pr((E \cap H) \cup (E \cap \overline{H})) = Pr(E \cap H) + Pr(E \cap \overline{H})$$

By the formula for conditional probability:
$$Pr(E) = [Pr(E|H) \times Pr(H)] + [Pr(E|\overline{H}) \times Pr(\overline{H})]$$

Substituting into Bayes’ Theorem, we get the second version:
$$Pr(H|E) = \frac{Pr(E|H) \times Pr(H)}{[Pr(E|H) \times Pr(H)] + [Pr(E|\overline{H}) \times Pr(\overline{H})]}$$

Here’s a simple example using this version of Bayes’ Theorem. There are two boxes, $A$ and $B$. $A$ contains 2 red balls and 8 yellow balls. $B$ contains 7 red balls and 3 yellow balls. You randomly select a box (you don’t know which) with equal probability and randomly select a ball from the box – it’s red. What is the probability that you have box $A$?

Let’s first label up the following sets:

$A$ Box $A$ was chosen
$B$ A red ball was picked

We want $Pr(A|R)$, which is:
$$Pr(R|A) \times Pr(A)$$
$$[Pr(R|A) \times Pr(A)] + [Pr(R|\overline{A}) \times Pr(\overline{A})]$$

We know the following values: $Pr(R|A) = \frac{2}{10}$; $Pr(R|\overline{A}) = \frac{7}{10}$; $Pr(A) = \frac{1}{2}$; $Pr(\overline{A}) = \frac{1}{2}$. So:
$$Pr(A|R) = \frac{\frac{2}{10} \times \frac{1}{2} + \frac{7}{10} \times \frac{1}{2}}{\frac{2}{10} \times \frac{1}{2} + \frac{7}{10} \times \frac{1}{2}} = \frac{2}{9}$$

Base rate fallacy
Here’s a more interesting example. Suppose that 85% of the taxis in Smallville are green and 15% are blue. There was an accident involving a taxi. A witness reports that the taxi was blue. Let’s say the witness correctly identifies the colour of taxis 80% of the time and incorrectly identifies them 20% of the time. How likely is it that the taxi was blue?

Let’s start again by labelling up some sets:

- **B** The taxi was blue
- **W** The witness said the taxi was blue

And we have the following values: \( Pr(W|B) = 80\%; \) \( Pr(B) = 15\%; \)
\( Pr(W|\overline{B}) = 20\%; \) \( Pr(\overline{B}) = 85\%. \) Now:

\[
Pr(B|W) = \frac{Pr(W|B) \times Pr(B)}{Pr(W|B) \times Pr(B) + Pr(W|\overline{B}) \times Pr(\overline{B})} = \frac{0.8 \times 0.15}{(0.8 \times 0.15) + (0.2 \times 0.85)} \approx 41\%
\]

So the witness is probably wrong. But this may surprise us: after all, the witness is 80% reliable. What’s gone wrong? We are ignoring the low base rate or initial probability that we’re dealing with a blue taxi. To ignore this is to commit a base rate fallacy.

The point is that, in spite of the reliability of our witness, the proportion of blue taxis in Smallville is only 15%. This means that the proportion of green taxis incorrectly identified as blue remains greater than the proportion of blue taxis correctly identified as blue. So, given the low base rate of blue taxis, it is still more likely that our very reliable witness misidentified a green taxi than correctly identified a blue taxi. It’s just so unlikely that they saw a blue car at all.

**Miracles**

The most famous application of Bayes’ Theorem is Hume’s treatment of miracles. Suppose that a witness reports that a man walked on water. Let’s call this \( T, \) for ‘testimony’. And let’s call the walking-on-water \( M, \) for ‘miracle’.

To calculate the probability that a miracle occurred, given that the witness reported it – \( Pr(M|T), \) we need \( Pr(M), Pr(T|M) \) and \( Pr(T|\overline{M}). \)

Let’s say that \( Pr(M) = \frac{1}{1M}, \) i.e. a miracle is highly unlikely.\(^5\) Let’s also say that \( Pr(T|M) = \frac{9}{10}, \) and \( Pr(T|\overline{M}) = \frac{1}{100}, \) i.e. the testimony is highly reliable. Now we can apply Bayes’ Theorem:

\[
Pr(M|T) = \frac{\frac{9}{10M} \times \frac{9}{10}}{\frac{9}{10M} \times \frac{9}{10} + \frac{1}{100M} \times \frac{1}{100}} \approx \frac{1}{10,000}
\]

\(^5\)Here, ‘1M’ is of course shorthand for ‘1,000,000’.
In spite of the reliability of the witness, it is still extremely unlikely that the miracle occurred. Let’s give the last word to Hume:

The plain consequence is (and it is a general maxim worthy of attention), “That no testimony is sufficient to establish a miracle, unless the testimony be of such a kind, that its falsehood would be more miraculous, than the fact, which it endeavours to establish; and even in that case there is a mutual destruction of arguments, and the superior only gives us an assurance suitable to that degree of force, which remains, after deducting the inferior.” (‘Of Miracles’, pp 115-6)

Degrees of belief

Let’s say that you leave your home with your umbrella and sunglasses. Do you think that it will rain? Well, you can’t be certain or you wouldn’t take your sunglasses. But then you can’t be certain that it will be sunny or you wouldn’t take your umbrella. Rather, you have a degree of belief that it will rain, and a degree of belief that it will be sunny, and we can represent these using probabilities. Of course, for most mental states this is the norm, since we aren’t certain of all that much.

How can we judge your degree of belief that some event will occur? An obvious suggestion is observation of your behaviour. Let’s say you go to the pub, you buy two pints and you take a seat at a table for two, checking your watch often. This all suggests that your degree of belief that someone will be joining you is high. At the extreme end of this thought, we could identify your belief with this behaviour. That would be a strong form of behaviourism about the mental.

That’s all quite imprecise, so let’s focus in on a subset of your behaviours: your willingness to accept betting odds. This approach to subjective probability was first offered by Frank Ramsey in his ‘Truth and Probability’ (1926). The approach is an intuitive one: to make your degrees of belief amenable to probabilistic treatment, we need to assign numerical values to your beliefs. And betting is an area where people often attach numerical values to their beliefs.

Betting

To understand this approach, it will be helpful to first have a reminder about betting odds, for those of you who haven’t had a flutter for a while. Our standard example will be the odds on offer at various bookies’ for Glastonbury 2019 headliners.

Skybet are offering 5/1 (pronounced ‘5 to 1’) that Paul McCartney will headline. What does this mean? If you bet £1 and you win, you will win £6 (5 × £1 + your initial stake of £1 back). Generally, if the odds are \( \frac{A}{B} \) and you win, you’ll win £\( \frac{A+B}{B} \) for every £1 bet. The higher this figure – \( \frac{A+B}{B} \) – the better the odds. If another bookies’ were to offer, say, 4/1 on McCartney’s headlining, they would be
McCartney’s odds of 5/1 are sometimes expressed as ‘5 to 1 against’, since the amount you would win is greater than the amount staked. Intuitively, he is unlikely to headline. I see that the odds for The Cure headlining are 1/2. Here, the second number is larger than the first. Odds such as this are sometimes expressed as ‘2 to 1 on’ since the amount you would win is less than the amount staked (here, for every £1 staked, you win 50p). Intuitively, The Cure are likely to headline. Sadly, Kendrick is at 33/1. Don’t bet on him.

Credences

The degree of confidence, or credence, that you have in event \( E \) is determined by the worst odds that you would accept, at least for small stakes. Let’s say I believe that Lady Gaga will headline Glastonbury. How confident am I? About 80%. If your credence in \( E \) is \( \frac{B}{A+B} \), then the worst odds you will be prepared to accept on \( E \) are \( A/B \). So the worst odds I would accept on Lady Gaga headlining are 1/4.

If a bookies’ were offering 1/10 on Lady Gaga headlining, I wouldn’t take the bet, since those are worse odds than 1/4. 1/10 represents a credence of about 90%, and my credence isn’t that high. If another bookies’ were offering 1/3 on Lady Gaga headlining, which represents a credence of 75%, I would take the bet, since my credence is higher.

This all makes intuitive sense: if you are certain that \( E \) will happen, you will accept any odds \( A/B \) (however small \( A \) is, as long as \( A \neq 0 \)). And if you are certain that \( E \) won’t happen, then you won’t accept any odds \( A/B \) (however small \( B \) is, as long as \( B \neq 0 \)).

Expected utility

Of course, my beliefs about Glastonbury headliners are peculiarly well-suited to this treatment. But, in principle, any action can be construed as a sort of gamble. You decided (wisely) to come to this lecture today. So presumably your degree of belief that you could get to the lecture was greater than your belief that the roads would be closed. Your degrees of belief in the lecture happening at this time and place were presumably also very high. You weren’t certain: this lecture could have been moved and you could have missed the message, but that’s unlikely.

In general, when someone performs an action, the expected utility of that action is greater than other actions. A very famous example of expected utility is Pascal’s wager. Let’s say that your credence in the existence of God is 1%. If you are correct that God doesn’t exist, you’ll take some pleasure in being correct. Let’s say this has positive utility 10. If God exists, though, and you don’t believe in Her, you’ll suffer eternal damnation. Let’s say this has negative utility 1M.

How about if you do believe in God? Well, if She exists, you’ll enjoy eternal happiness. Let’s say this has positive utility 1M. And if you believe and She
doesn’t exists, you’ll take some displeasure in being wrong, say negative utility 10.

We can helpfully display this data in a table:

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>+1M</td>
<td>-1M</td>
</tr>
<tr>
<td>G</td>
<td>-10</td>
<td>+10</td>
</tr>
</tbody>
</table>

And we’ve said that \( Pr(G) = 0.01 \) and \( Pr(\overline{G}) = 0.99 \). The expected utility of believing in God is \( (1M \times 0.01) + (-10 \times 0.99) = 9,990.1 \). The expected utility of not believing is \( (-1M \times 0.01) + (10 \times 0.99) = -9990.1 \). So you should believe in God.

**Dutch books**

If we interpret the probability calculus subjectively, using betting odds, then the Kolmogorov axioms have a rational compulsion. It turns out that, if your reasoning fails to conform to any one of them, you will accept a bet that you are guaranteed to lose. To see this we need to first discuss the notion of a Dutch book, which we’ll introduce by example.

Let’s say it’s the final of Rupaul’s Drag Race and one of three queens will win: Alaska, Roxxy and Jinkx. You discover that Ladbroke’s is offering 3/1 on Alaska, 3/1 on Roxxy and 3/1 on Jinkx. What should you do? You should take every penny you have and put \( \frac{1}{3} \) of your money on each: you are bound to win. Why?

Let’s say you start with £3 and put £1 on each. If Alaska wins, you’ll get £4 back; if Roxxy wins, you’ll get £4 back; if Jinkx wins, you’ll get £4 back. And one of them will win so, in any case, you end up with more than you started.

The bookies is bound to lose. Of course, no actual bookies would set up such odds and guarantee that they lose money. (Although, apparently, such situations can arise in international sport.) If you accept betting odds that guarantee that you’ll lose money, we say that you’re open to a Dutch book. In this example, the bookies is open to a Dutch book. But of course bettors themselves can also be open to Dutch books, if they accept the wrong odds.

If our credences fail to conform to the axioms of probability theory, we leave ourselves open to Dutch books. Let’s say I have a credence of 60% that it will snow tomorrow. And I also have a credence of 60% that it won’t snow tomorrow. This plainly violates the probabilistic theorem we proved that \( Pr(X) + Pr(\overline{X}) = 1 \). And I am open to a Dutch book: 60% credence shows that I’d accept odds of 2/3 that it will snow tomorrow and odds of 2/3 that it won’t.

Let’s say that I start out with £6 and I put £3 on each outcome. It will either snow or not snow, so I’ll lose one of the bets and win the other. The one I lose will cost me £3. The one I win will give me £5. So I end up with £5 – overall, I’ve lost £1. A Dutch book can be made against you if, and only if, your credences fail to conform to the axioms of probability.
This is all very exciting but it’s important to note some of the idealisations being made here. We are assuming that, for any person \( S \) at any time \( t \) and for any outcome \( E \), there will be a number between 0 and 1 that represents \( S \)’s credence in \( E \) at \( t \). There are obvious counterexamples to be had: some outcomes we obviously haven’t entertained. But, even in the cases we have considered, is it plausible that we can quantify credences in this way? Or are the idealisations involved so great that they shed little light on actual human behaviour? Questions for another time but, if you’re interested, I recommend starting with Ramsey’s ‘Truth and Probability’.