Sometimes when we design DFAs we have quite a bit of freedom, and it is possible to have two different DFAs for the same task. Which one of these DFAs is the"best" one? One answer to this question is to try to build a *minimal DFA* — one with the smallest possible number of states.

But building a minimal DFA from scratch can be difficult. Instead, we will show a procedure that converts any DFA into a minimal one. You will also learn how to tell if a DFA is indeed minimal (that is, how to say for sure that you cannot make it any smaller).

## 1 Minimal DFAs, reachable states, distinguishable states

We first need to define the notion of a "minimal DFA" more formally. Intuitively, a DFA is minimal if no DFA with fewer states does the same task; so the languages of the two DFAs must be different.

**Definition 1.** Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA. Then $M$ is *minimal* if for every DFA $M' = (Q', \Sigma, \delta', q_0', F')$ such that $|Q'| < |Q|$, $L(M') \neq L(M)$.

Now let's imagine that we have a minimal DFA and try to see what it must look like. First, it should be clear that in a minimal $M$ every state $q$ must be "reachable": There must be some way to get to it from the start state. If there is no path that leads from the start state to $q$, then we can eliminate $q$ (since nothing leads to it) together with all its outgoing transitions to get a smaller DFA.

But it turns out that there is another feature that minimal DFAs must have. To understand this we need to look not at single states but at pairs of states. Suppose you are looking at two different states $q$ and $q'$ in your DFA. Now consider what happens if you run your DFA "in parallel" starting at $q$ and $q'$ on the same input $w$. Several different things can happen: After reading $w$, both runs could end in accepting states, both could end in rejecting states, or one could end in an accepting state, while the other ends in a rejecting state. In the last case the two states are called "distinguishable".

Formally, here is how we define "reachable" and "distinguishable".

**Definition 2.** Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA. A states $q$ is *reachable* in $M$ if there exists a string $w = w_1 w_2 \dots w_m$ over the alphabet $\Sigma$ and a sequence of states $q_1, \dots, q_m$ such that

1. $q_{i+1} = \delta(q_i, w_i)$, for $i = 0, \dots, m-1$, and
2. $q_m = q$

The string $w$ is said to *reach* state $q$ in $M$.

**Definition 3.** Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA. A pair of states $q, q' \in Q$ is *distinguishable* in $M$ if there exists a string $w = w_1 w_2 \dots w_m$ over the alphabet $\Sigma$ and two sequences of states $q_1, \dots, q_m$ and $q_1', \dots, q_m'$ in $Q$ such that

1. $q_1 = q$ and $q'_1 = q'$,
2. $q_{i+1} = \delta(q_i, w_i)$ and $q'_{i+1} = \delta(q'_i, w_i)$, for $i = 1, \ldots, m-1$, and
3. Exactly one of the states $q_m$ and $q'_m$ is in $F$.

## 2 Characterizing minimal DFAs

We now describe the main feature of a minimal DFA.

**Theorem 4.** *Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA. Then $M$ is minimal if and only if every state of $M$ is reachable and every pair of states of $M$ is distinguishable.*

To prove this theorem, we have to do two things: First, we have to show that if every state is reachable and every pair of states of $M$ is distinguishable, then $M$ is minimal. Then, we have to show that if $M$ is minimal, then every state is reachable and every pair of states is distinguishable.

**Lemma 5.** *If every state of $M$ is reachable and every pair of states of $M$ is distinguishable, then $M$ is minimal.*

*Proof sketch.* We need to argue that if every state of $M$ is reachable and every pair of states of $M$ is distinguishable, then $M$ is minimal, that is there can be no smaller $M' = (Q', \Sigma, \delta', q'_0, F')$ (i.e., one with $|Q'| < |Q|$) for the same language. We will assume such an $M'$ exists and argue that something must be wrong with it.

Because every state of $M$ is reachable, for every state $q_i \in Q$, there is a string $w_i$ that reaches $q_i$ in $M$. By the pigeonhole principle, there exists a reachable state $q'$ of $Q'$ and a pair of strings $w_i, w_j$ (where $w_i \neq w_j$) such that $w_i$ reaches $q'$ in $M'$ and $w_j$ reaches $q'$ in $M'$. By assumption, $q_i$ and $q_j$ are distinguishable in $M$, so there exists a string $w$ that distinguishes them. Now think what happens when we run $M$ on the strings $w_i w$ and $w_j w$. Since $q_i$ and $q_j$ are distinguishable, exactly one of $w_i w$ and $w_j w$ reaches an accepting state of $M$. On the other hand, both $w_i w$ and $w_j w$ reach the same state of $M'$. Therefore the languages of $M$ and $M'$ must be different. □

For the other direction, we need to argue that if $M$ is minimal, then every state of $M$ is reachable and every pair of states of $M$ is distinguishable. We will reason by contradiction: If either $M$ has a state that is not reachable, or $M$ has a pair of states that are indistinguishable, then we will show that $M$ is not minimal – it can be made smaller. Let us start by seeing what happens when $M$ has an unreachable state.

**Lemma 6.** *If $M$ has an unreachable state, then $M$ is not minimal.*

*Proof.* Suppose $M$ has a state $q$ that is not reachable. Let $M'$ be the DFA obtained by removing *all* unreachable states from $M$, together with their outgoing transitions. That is $M' = (Q', \Sigma, \delta', q_0, F')$ where

1. The states of $M'$ are: $Q' = \{q : q \in Q \text{ and } q \text{ is reachable in } M\}$,

2. The transitions of $M'$ are: $\delta'(q, a) = \delta(q, a)$ for every reachable $q$ and every $a \in \Sigma$, and
3. The accepting states of $M'$ are: $F' = \{q : q \in F \text{ and } q \text{ is reachable in } M\}$

Then $|Q'| < |Q|$ (because $M$ has a state that is not reachable), but $L(M') = L(M)$, so $M'$ is not minimal. □

Now we are left with the case when $M$ has a pair of states that are indistinguishable.

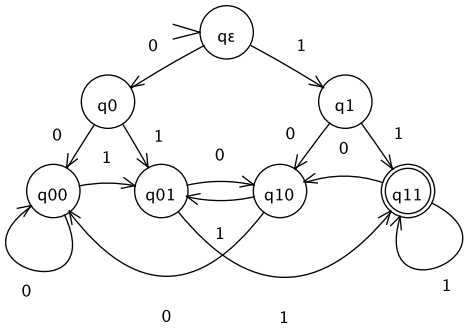
**Lemma 7.** *If $M$ has a pair of indistinguishable states, then $M$ is not minimal.*

We will explain and prove this lemma in the following section.

## 3 Merging indistinguishable states

The way we will argue this is by showing that pairs of indistinguishable states can be merged together. This merging operation will reduce the number of states. But we must be careful: What if after merging the states we change the behavior of the DFA? And how can we even speak about merging two states whose transitions go out to different places? What should we do with the transitions out of the merged state?

To show how this merging can be performed consitently, let's look at an example.



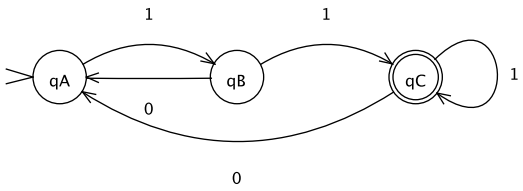

In this DFA, the following pairs of states are indistinguishable:

$$(q_\varepsilon, q_0) \quad (q_\varepsilon, q_{00}) \quad (q_\varepsilon, q_{10}) \quad (q_0, q_{00}) \quad (q_0, q_{10}) \quad (q_{00}, q_{10}) \quad (q_1, q_{01}) \tag{1}$$

(In addition, every state is indistinguishable from itself.) These states can be split into classes: Class $A$, consisting of $\{q_\varepsilon, q_0, q_{00}, q_{10}\}$, and class $B$ consisting of $\{q_1, q_{01}\}$. State $q_{11}$ is only indistinguishable from itself, so we put it in a class $C$ consisting of $\{q_{11}\}$ only.

Now observe that a remarkable thing happens with the transitions: The transitions among groups of states are consistent. For instance, on a 0-transition, every state from class $A$ moves back to

class $A$, and on a 1-transition, every state from class $A$ moves to class $B$. The same thing happens with the other classes. So we can take together all the states from the same class, merge them together into new "megastates", and obtain a smaller DFA:




In fact, this merging of indistinguishable states into "megastates" can always be done in such a way. The reason is that if a pair of states $q_i$ and $q_j$ are indistinguishable, then for every $a \in \Sigma$, $r_i = \delta(q_i, a)$ and $r_j = \delta(q_j, a)$ must also be indistinguishable: For if $r_i$ and $r_j$ can be distinguished by $w$ , then $q_i$ and $q_j$ can be distinguished by $aw$. So transitions out of indistinguishable states point into indistinguishable states.

With this in mind, we can prove Lemma 7.

*Proof of Lemma 7.* Assume $M$ has a pair of distinguishable states. We will show how to construct an equivalent DFA $M' = (Q', \Sigma, \delta', q'_0, F')$ with fewer states.

To do so, we divide the states $Q$ of $M$ into *indistinguishability classes* $q'_0, q'_1, \ldots, q'_{m'}$. Each of these classes represents a subset of states of $M$ which are mutually indistinguishable. We will say that a state $q$ is *represented* by its indistinguishability class $q'$. We can assume that $q_0$ is represented by $q'_0$. Since $M$ has at least two indistinguishable states, it follows that $m' < |Q|$.

We can now describe $M'$:

1. States: $Q' = \{q'_0, q'_1, \ldots, q'_{m'}\}$,
2. Transitions: For every $q' \in Q'$ and $a \in \Sigma$, if $q'$ represents $q$, then $\delta'(q', a)$ represents $\delta(q, a)$.
3. Accepting states: $F' = \{q' : \text{The states in class } q' \text{ are all in } F\}$.

For this definition to make sense, we have to argue that the transitions of $M'$ are well defined: If $q_i$ and $q_j$ are both represented by the same state, we need to know that $r_i = \delta(q_i, a)$ and $r_j = \delta(q_j, a)$ are both represented by the same state. Indeed, this must be the case: If $r_i$ and $r_j$ are represented by different states, then they must be distinguishable — say by $w$. But then $q_i$ and $q_j$ are distinguishable by $aw$, and this contradicts the fact that they belong to the same indistinguishability class.

Now we can argue that $L(M) = L(M')$. Let $w = w_1 \ldots w_k$ be any string. Suppose that when we run $M$ on input $w$, $M$ goes through the states $q_0, q_1, \ldots, q_k$, that is, $\delta(q_i, w_{i+1}) = q_{i+1}$ for $i = 0, \ldots, k-1$. Then on the same input $w$, $M'$ will go through some sequence of states $q'_0, q'_1, \ldots, q'_k$, where $q'_i$ is the representative of $q_i$.

The last state $q'_k$ will be accepting if and only if $q_k$ is accepting: If $q_k$ is rejecting, then $q'_k$ contains a rejecting state so it is also rejecting. On the other hand, if $q_k$ is accepting, then so must be all the other states represented by $q'_k$ (because they are indistinguishable), so $q'_k$ is also accepting.

It follows that $M$ accepts $w$ *if and only if* $M'$ accepts $w$ – so $M$ and $M'$ are equivalent. But $M'$ has fewer states than $M$. □

# 4 The DFA minimization algorithm

Now we understand what minimal DFAs must look like: All their states are reachable, and all their pairs of states are distinguishable. Moreover, we saw that if $M$ is not minimal, we can make it smaller by grouping together states into indistinguishability classes. It remains to see how we can find these indistinguishability classes systematically.

To do so, we must first figure out which pairs of states of $M$ are indistinguishable. An easier task is to find the pairs that are distinguishable. To do so, we will iteratively update a "table" $X$ of (unordered) pairs of distinguishable states $(q, q')$ using the following rules:

> **Initialization**: Remove all unreachable states of $M$. Set $X$ to be empty.
>
> **Rule 1:** If $q$ is accepting and $q'$ is rejecting, add the pair $(q, q')$ to $X$.
>
> **Rule 2:** If $(q, q')$ is already in $X$ and $r, r'$ is a pair such that $q = \delta(r, a)$ and $q' = \delta(r', a)$ for some $a \in \Sigma$, then add the pair $(r, r')$ to $X$.

We apply rules 1 and 2 as long as new pairs can be added to $X$ using these rules. Once we are finished, $X$ will contain all pairs of distinguishable states.

All pairs of unmarked states will be indistinguishable and they can be merged together into indistinguishable classes. The resulting DFA will be minimal.

Let us explain how this works on the above example. We start with an empty $X$. Applying Rule 1, we can add all the pairs $(q_\varepsilon, q_{11}), (q_0, q_{11}), (q_1, q_{11}), (q_{00}, q_{11}), (q_{01}, q_{11})$, and $(q_{10}, q_{11})$ to $X$. After this step,

$$X = \{(q_\varepsilon, q_{11}), (q_0, q_{11}), (q_1, q_{11}), (q_{00}, q_{11}), (q_{01}, q_{11}), (q_{10}, q_{11})\}.$$

Now we can start applying Rule 2. For example, $(q_1, q_{11})$ is already in $X$, and we see that $\delta(q_\varepsilon, 1) = q_1$, and $\delta(q_1, 1) = q_{11}$, so we also add the pair $(q_\varepsilon, q_1)$ to $X$. Similarly, we can also include the pairs $(q_0, q_1)$, $(q_{00}, q_1)$, $(q_{10}, q_1)$, $(q_\varepsilon, q_{01})$, $(q_0, q_{01})$, $(q_{00}, q_{01})$, and $(q_{10}, q_{01})$. Now

$$\begin{aligned} X = \{&(q_\varepsilon, q_{11}), (q_0, q_{11}), (q_1, q_{11}), (q_{00}, q_{11}), (q_{01}, q_{11}), (q_{10}, q_{11}), \\ &(q_\varepsilon, q_1), (q_0, q_1), (q_{00}, q_1), (q_{10}, q_1), (q_\varepsilon, q_{01}), (q_0, q_{01}), (q_{00}, q_{01}), (q_{10}, q_{01})\}. \end{aligned}$$

At this point, there is nothing more to add using Rules 1 and 2. The only pairs of states not included in $X$ are now those in the list (1). Those are the pairs of indistinguishable states.

After merging together the indistinguishable states, we obtain the minimized DFA.