

## Week4 wednesday

Regular sets are not the end of the story

- Many nice / simple / important sets are not regular
- Limitation of the finite-state automaton model: Can't "count", Can only remember finitely far into the past, Can't backtrack, Must make decisions in "real-time"
- We know actual computers are more powerful than this model...

The **next** model of computation. Idea: allow some memory of unbounded size. How?

- To generalize regular expressions: **context-free grammars**
- To generalize NFA: **Pushdown automata**, which is like an NFA with access to a stack: Number of states is fixed, number of entries in stack is unbounded. At each step (1) Transition to new state based on current state, letter read, and top letter of stack, then (2) (Possibly) push or pop a letter to (or from) top of stack. Accept a string iff there is some sequence of states and some sequence of stack contents which helps the PDA process the entire input string and ends in an accepting state.

Is there a PDA that recognizes the nonregular language  $\{0^n 1^n \mid n \geq 0\}$ ?



The PDA with state diagram above can be informally described as:

Read symbols from the input. As each 0 is read, push it onto the stack. As soon as 1s are seen, pop a 0 off the stack for each 1 read. If the stack becomes empty and we are at the end of the input string, accept the input. If the stack becomes empty and there are 1s left to read, or if 1s are finished while the stack still contains 0s, or if any 0s appear in the string following 1s, reject the input.

Trace the computation of this PDA on the input string 01.

Trace the computation of this PDA on the input string 011.

A PDA recognizing the set {  $0^n 1^n$  } can be informally described as:

Read symbols from the input. As each 0 is read, push it onto the stack. As soon as 1s are seen, pop a 0 off the stack for each 1 read. If the stack becomes empty and there is exactly one 1 left to read, read that 1 and accept the input. If the stack becomes empty and there are either zero or more than one 1s left to read, or if the 1s are finished while the stack still contains 0s, or if any 0s appear in the input following 1s, reject the input.

Modify the state diagram below to get a PDA that implements this description:



**Definition** A **pushdown automaton** (PDA) is specified by a 6-tuple  $(Q, \Sigma, \Gamma, \delta, q_0, F)$  where  $Q$  is the finite set of states,  $\Sigma$  is the input alphabet,  $\Gamma$  is the stack alphabet,

$$\delta : Q \times \Sigma_\epsilon \times \Gamma_\epsilon \rightarrow \mathcal{P}(Q \times \Gamma_\epsilon)$$

is the transition function,  $q_0 \in Q$  is the start state,  $F \subseteq Q$  is the set of accept states.

Draw the state diagram and give the formal definition of a PDA with  $\Sigma = \Gamma$ .

Draw the state diagram and give the formal definition of a PDA with  $\Sigma \cap \Gamma = \emptyset$ .

*Extra practice:* Consider the state diagram of a PDA with input alphabet  $\Sigma$  and stack alphabet  $\Gamma$ .

Label	means
$a, b; c$ when $a \in \Sigma, b \in \Gamma, c \in \Gamma$	
$a, \varepsilon; c$ when $a \in \Sigma, c \in \Gamma$	
$a, b; \varepsilon$ when $a \in \Sigma, b \in \Gamma$	
$a, \varepsilon; \varepsilon$ when $a \in \Sigma$	

How does the meaning change if  $a$  is replaced by  $\varepsilon$ ?

*Note: alternate notation is to replace ; with  $\rightarrow$*

## Week4 friday

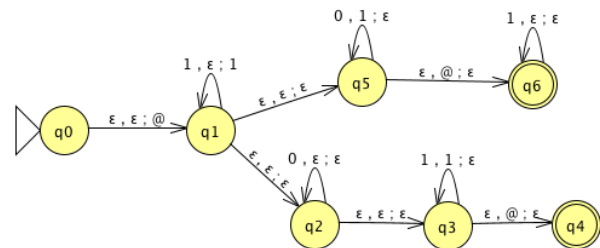
For the PDA state diagrams below,  $\Sigma = \{0, 1\}$ .

Mathematical description of language

State diagram of PDA recognizing language  
 $\Gamma = \{\$, \#\}$



$\Gamma = \{ @, 1 \}$



$$\{0^i 1^j 0^k \mid i, j, k \geq 0\}$$

*Big picture:* PDAs were motivated by wanting to add some memory of unbounded size to NFA. How do we accomplish a similar enhancement of regular expressions to get a syntactic model that is more expressive?

DFA, NFA, PDA: Machines process one input string at a time; the computation of a machine on its input string reads the input from left to right.

Regular expressions: Syntactic descriptions of all strings that match a particular pattern; the language described by a regular expression is built up recursively according to the expression's syntax

**Context-free grammars:** Rules to produce one string at a time, adding characters from the middle, beginning, or end of the final string as the derivation proceeds.

Term	Typical symbol	Definition
<b>Context-free grammar</b> (CFG)	$G$	$G = (V, \Sigma, R, S)$
<b>Variables</b>	$V$	Finite set of symbols that represent phases in production pattern
<b>Terminals</b>	$\Sigma$	Alphabet of symbols of strings generated by CFG $V \cap \Sigma = \emptyset$
<b>Rules</b>	$R$	Each rule is $A \rightarrow u$ with $A \in V$ and $u \in (V \cup \Sigma)^*$
Start variable	$S$	Usually on LHS of first / topmost rule
<b>Derivation</b>	$S \Rightarrow \dots \Rightarrow w$	Sequence of substitutions in a CFG Start with start variable, apply one rule to one occurrence of a variable at a time
<b>Language</b> generated by the CFG $G$	$L(G)$	$\{w \in \Sigma^* \mid \text{there is derivation in } G \text{ that ends in } w\} = \{w \in \Sigma^* \mid S \Rightarrow^* w\}$
<b>Context-free language</b>		A language that is the language generated by some CFG
Sipser pages 102-103		

**Examples of context-free grammars, derivations in those grammars, and the languages generated by those grammars**

$G_1 = (\{S\}, \{0\}, R, S)$  with rules

$$S \rightarrow 0S$$

$$S \rightarrow 0$$

In  $L(G_1)$  ...

Not in  $L(G_1)$  ...

$$G_2 = (\{S\}, \{0, 1\}, R, S)$$

$$S \rightarrow 0S \mid 1S \mid \varepsilon$$

In  $L(G_2) \dots$

Not in  $L(G_2) \dots$

$(\{S, T\}, \{0, 1\}, R, S)$  with rules

$$S \rightarrow T1T1T1T$$

$$T \rightarrow 0T \mid 1T \mid \varepsilon$$

In  $L(G_3) \dots$

Not in  $L(G_3) \dots$

$G_4 = (\{A, B\}, \{0, 1\}, R, A)$  with rules

$$A \rightarrow 0A0 \mid 0A1 \mid 1A0 \mid 1A1 \mid 1$$

In  $L(G_4)$  ...

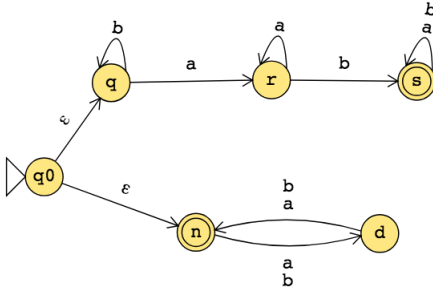
Not in  $L(G_4)$  ...

*Extra practice:* Is there a CFG  $G$  with  $L(G) = \emptyset$ ?



## Week3 monday

The state diagram of an NFA over  $\{a, b\}$  is below. The formal definition of this NFA is:



The language recognized by this NFA is:

Suppose  $A_1, A_2$  are languages over an alphabet  $\Sigma$ . **Claim:** if there is a NFA  $N_1$  such that  $L(N_1) = A_1$  and NFA  $N_2$  such that  $L(N_2) = A_2$ , then there is another NFA, let's call it  $N$ , such that  $L(N) = A_1 \cup A_2$ .

**Proof idea:** Use nondeterminism to choose which of  $N_1, N_2$  to run.

**Formal construction:** Let  $N_1 = (Q_1, \Sigma, \delta_1, q_1, F_1)$  and  $N_2 = (Q_2, \Sigma, \delta_2, q_2, F_2)$  and assume  $Q_1 \cap Q_2 = \emptyset$  and that  $q_0 \notin Q_1 \cup Q_2$ . Construct  $N = (Q, \Sigma, \delta, q_0, F_1 \cup F_2)$  where

- $Q =$
- $\delta : Q \times \Sigma_\epsilon \rightarrow \mathcal{P}(Q)$  is defined by, for  $q \in Q$  and  $a \in \Sigma_\epsilon$ :

*Proof of correctness would prove that  $L(N) = A_1 \cup A_2$  by considering an arbitrary string accepted by  $N$ , tracing an accepting computation of  $N$  on it, and using that trace to prove the string is in at least one of  $A_1, A_2$ ; then, taking an arbitrary string in  $A_1 \cup A_2$  and proving that it is accepted by  $N$ . Details left for extra practice.*

Over the alphabet  $\{a, b\}$ , the language  $L$  described by the regular expression  $\Sigma^*a\Sigma^*b$

includes the strings

and excludes the strings

The state diagram of a NFA recognizing  $L$  is:

Suppose  $A_1, A_2$  are languages over an alphabet  $\Sigma$ . **Claim:** if there is a NFA  $N_1$  such that  $L(N_1) = A_1$  and NFA  $N_2$  such that  $L(N_2) = A_2$ , then there is another NFA, let's call it  $N$ , such that  $L(N) = A_1 \circ A_2$ .

**Proof idea:** Allow computation to move between  $N_1$  and  $N_2$  “spontaneously” when reach an accepting state of  $N_1$ , guessing that we’ve reached the point where the two parts of the string in the set-wise concatenation are glued together.

**Formal construction:** Let  $N_1 = (Q_1, \Sigma, \delta_1, q_1, F_1)$  and  $N_2 = (Q_2, \Sigma, \delta_2, q_2, F_2)$  and assume  $Q_1 \cap Q_2 = \emptyset$ . Construct  $N = (Q, \Sigma, \delta, q_0, F)$  where

- $Q =$
- $q_0 =$
- $F =$
- $\delta : Q \times \Sigma_\varepsilon \rightarrow \mathcal{P}(Q)$  is defined by, for  $q \in Q$  and  $a \in \Sigma_\varepsilon$ :

$$\delta((q, a)) = \begin{cases} \delta_1((q, a)) & \text{if } q \in Q_1 \text{ and } q \notin F_1 \\ \delta_1((q, a)) & \text{if } q \in F_1 \text{ and } a \in \Sigma \\ \delta_1((q, a)) \cup \{q_2\} & \text{if } q \in F_1 \text{ and } a = \varepsilon \\ \delta_2((q, a)) & \text{if } q \in Q_2 \end{cases}$$

*Proof of correctness would prove that  $L(N) = A_1 \circ A_2$  by considering an arbitrary string accepted by  $N$ , tracing an accepting computation of  $N$  on it, and using that trace to prove the string can be written as the result of concatenating two strings, the first in  $A_1$  and the second in  $A_2$ ; then, taking an arbitrary string in  $A_1 \circ A_2$  and proving that it is accepted by  $N$ . Details left for extra practice.*

Suppose  $A$  is a language over an alphabet  $\Sigma$ . **Claim:** if there is a NFA  $N$  such that  $L(N) = A$ , then there is another NFA, let's call it  $N'$ , such that  $L(N') = A^*$ .

**Proof idea:** Add a fresh start state, which is an accept state. Add spontaneous moves from each (old) accept state to the old start state.

**Formal construction:** Let  $N = (Q, \Sigma, \delta, q_1, F)$  and assume  $q_0 \notin Q$ . Construct  $N' = (Q', \Sigma, \delta', q_0, F')$  where

- $Q' = Q \cup \{q_0\}$
- $F' = F \cup \{q_0\}$
- $\delta' : Q' \times \Sigma_\varepsilon \rightarrow \mathcal{P}(Q')$  is defined by, for  $q \in Q'$  and  $a \in \Sigma_\varepsilon$ :

$$\delta'((q, a)) = \begin{cases} \delta((q, a)) & \text{if } q \in Q \text{ and } q \notin F \\ \delta((q, a)) & \text{if } q \in F \text{ and } a \in \Sigma \\ \delta((q, a)) \cup \{q_1\} & \text{if } q \in F \text{ and } a = \varepsilon \\ \{q_1\} & \text{if } q = q_0 \text{ and } a = \varepsilon \\ \emptyset & \text{if } q = q_0 \text{ and } a \in \Sigma \end{cases}$$

*Proof of correctness would prove that  $L(N') = A^*$  by considering an arbitrary string accepted by  $N'$ , tracing an accepting computation of  $N'$  on it, and using that trace to prove the string can be written as the result of concatenating some number of strings, each of which is in  $A$ ; then, taking an arbitrary string in  $A^*$  and proving that it is accepted by  $N'$ . Details left for extra practice.*

**Application:** A state diagram for a NFA over  $\Sigma = \{a, b\}$  that recognizes  $L((\Sigma^*b)^*)$ :

**True or False:** The state diagram of any DFA is also the state diagram of a NFA.

**True or False:** The state diagram of any NFA is also the state diagram of a DFA.

**True or False:** The formal definition  $(Q, \Sigma, \delta, q_0, F)$  of any DFA is also the formal definition of a NFA.

**True or False:** The formal definition  $(Q, \Sigma, \delta, q_0, F)$  of any NFA is also the formal definition of a DFA.

## Week3 wednesday

Consider the state diagram of an NFA over  $\{a, b\}$ :



The language recognized by this NFA is

The state diagram of a DFA recognizing this same language is:

Suppose  $A$  is a language over an alphabet  $\Sigma$ . **Claim:** if there is a NFA  $N$  such that  $L(N) = A$  then there is a DFA  $M$  such that  $L(M) = A$ .

**Proof idea:** States in  $M$  are “macro-states” – collections of states from  $N$  – that represent the set of possible states a computation of  $N$  might be in.

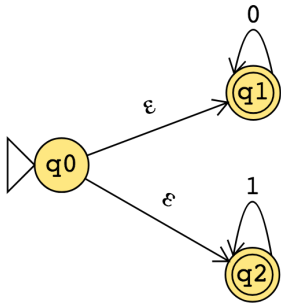
**Formal construction:** Let  $N = (Q, \Sigma, \delta, q_0, F)$ . Define

$$M = ( \mathcal{P}(Q), \Sigma, \delta', q', \{X \subseteq Q \mid X \cap F \neq \emptyset\} )$$

where  $q' = \{q \in Q \mid q = q_0 \text{ or is accessible from } q_0 \text{ by spontaneous moves in } N\}$  and

$\delta'((X, x)) = \{q \in Q \mid q \in \delta(r, x) \text{ for some } r \in X \text{ or is accessible from such an } r \text{ by spontaneous moves in } N\}$

Consider the state diagram of an NFA over  $\{0, 1\}$ . Use the “macro-state” construction to find an equivalent DFA.



Prune this diagram to get an equivalent DFA with only the “macro-states” reachable from the start state.

Suppose  $A$  is a language over an alphabet  $\Sigma$ . **Claim:** if there is a regular expression  $R$  such that  $L(R) = A$ , then there is a NFA, let's call it  $N$ , such that  $L(N) = A$ .

**Structural induction:** Regular expression is built from basis regular expressions using inductive steps (union, concatenation, Kleene star symbols). Use constructions to mirror these in NFAs.

**Application:** A state diagram for a NFA over  $\{a, b\}$  that recognizes  $L(a^*(ab)^*)$ :

Suppose  $A$  is a language over an alphabet  $\Sigma$ . **Claim:** if there is a DFA  $M$  such that  $L(M) = A$ , then there is a regular expression, let's call it  $R$ , such that  $L(R) = A$ .

**Proof idea:** Trace all possible paths from start state to accept state. Express labels of these paths as regular expressions, and union them all.

1. Add new start state with  $\varepsilon$  arrow to old start state.
2. Add new accept state with  $\varepsilon$  arrow from old accept states. Make old accept states non-accept.
3. Remove one (of the old) states at a time: modify regular expressions on arrows that went through removed state to restore language recognized by machine.

**Application:** Find a regular expression describing the language recognized by the DFA with state diagram



**Conclusion:** For each language  $L$ ,

**There is a DFA that recognizes  $L$  if and only if**

**There is a NFA that recognizes  $L$  if and only if**

**There is a regular expression that describes  $L$**

A language is called **regular** when any (hence all) of the above three conditions are met.

## Week2 monday

**Review:** Formal definition of DFA:  $M = (Q, \Sigma, \delta, q_0, F)$

- Finite set of states  $Q$
- Alphabet  $\Sigma$
- Transition function  $\delta$
- Start state  $q_0$
- Accept (final) states  $F$

In the state diagram of  $M$ , how many outgoing arrows are there from each state?

$M = (\{q, r, s\}, \{a, b\}, \delta, q, \{s\})$  where  $\delta$  is (rows labelled by states and columns labelled by symbols):

$\delta$	$a$	$b$
$q$	$r$	$q$
$r$	$r$	$s$
$s$	$s$	$s$

The state diagram for  $M$  is

Give two examples of strings that are accepted by  $M$  and two examples of strings that are rejected by  $M$ :

Add “labels” for states in the state diagram, e.g. “have not seen any of desired pattern yet” or “sink state”.

We can use the analysis of the roles of the states in the state diagram to describe the language recognized by the DFA.

$L(M) =$

A regular expression describing  $L(M)$  is

Let the alphabet be  $\Sigma_1 = \{0, 1\}$ .

A state diagram for a DFA that recognizes  $\{w \mid w \text{ contains at most two 1's}\}$  is

A state diagram for a DFA that recognizes  $\{w \mid w \text{ contains more than two 1's}\}$  is



*Extra example:* A state diagram for DFA recognizing

$$\{w \mid w \text{ is a string over } \{0,1\} \text{ whose length is not a multiple of } 3\}$$

Let  $n$  be an arbitrary positive integer. What is a formal definition for a DFA recognizing

$$\{w \mid w \text{ is a string over } \{0,1\} \text{ whose length is not a multiple of } n\}?$$

## Week2 wednesday

Suppose  $A$  is a language over an alphabet  $\Sigma$ . By definition, this means  $A$  is a subset of  $\Sigma^*$ . **Claim:** if there is a DFA  $M$  such that  $L(M) = A$  then there is another DFA, let's call it  $M'$ , such that  $L(M') = \overline{A}$ , the complement of  $A$ , defined as  $\{w \in \Sigma^* \mid w \notin A\}$ .

**Proof idea:**

**Proof:**

A useful (optional) bit of terminology: the **iterated transition function** of a DFA  $M = (Q, \Sigma, \delta, q_0, F)$  is defined recursively by

$$\delta^*(q, w) = \begin{cases} q & \text{if } q \in Q, w = \varepsilon \\ \delta(q, a) & \text{if } q \in Q, w = a \in \Sigma \\ \delta(\delta^*(q, u), a) & \text{if } q \in Q, w = ua \text{ where } u \in \Sigma^* \text{ and } a \in \Sigma \end{cases}$$

Using this terminology,  $M$  accepts a string  $w$  over  $\Sigma$  if and only if  $\delta^*(q_0, w) \in F$ .

Fix  $\Sigma = \{a, b\}$ . A state diagram for a DFA that recognizes  $\{w \mid w \text{ has } ab \text{ as a substring and is of even length}\}$ :

Suppose  $A_1, A_2$  are languages over an alphabet  $\Sigma$ . **Claim:** if there is a DFA  $M_1$  such that  $L(M_1) = A_1$  and DFA  $M_2$  such that  $L(M_2) = A_2$ , then there is another DFA, let's call it  $M$ , such that  $L(M) = A_1 \cap A_2$ .

**Proof idea:**

**Formal construction:**

**Application:** When  $A_1 = \{w \mid w \text{ has } ab \text{ as a substring}\}$  and  $A_2 = \{w \mid w \text{ is of even length}\}$ .

Suppose  $A_1, A_2$  are languages over an alphabet  $\Sigma$ . **Claim:** if there is a DFA  $M_1$  such that  $L(M_1) = A_1$  and DFA  $M_2$  such that  $L(M_2) = A_2$ , then there is another DFA, let's call it  $M$ , such that  $L(M) = A_1 \cup A_2$ .  
*Sipser Theorem 1.25, page 45*

**Proof idea:**

**Formal construction:**

**Application:** A state diagram for a DFA that recognizes  $\{w \mid w \text{ has } ab \text{ as a substring or is of even length}\}$ :

## Week2 friday

<b>Nondeterministic finite automaton</b> $M = (Q, \Sigma, \delta, q_0, F)$	
Finite set of states $Q$	Can be labelled by any collection of distinct names. Default: $q_0, q_1, \dots$
Alphabet $\Sigma$	Each input to the automaton is a string over $\Sigma$ .
Arrow labels $\Sigma_\varepsilon$	$\Sigma_\varepsilon = \Sigma \cup \{\varepsilon\}$ .
Transition function $\delta$	Arrows in the state diagram are labelled either by symbols from $\Sigma$ or by $\varepsilon$ $\delta : Q \times \Sigma_\varepsilon \rightarrow \mathcal{P}(Q)$ gives the <b>set of possible next states</b> for a transition from the current state upon reading a symbol or spontaneously moving.
Start state $q_0$	Element of $Q$ . Each computation of the machine starts at the start state.
Accept (final) states $F$	$F \subseteq Q$ .
$M$ accepts the input string	if and only if <b>there is</b> a computation of $M$ on the input string that processes the whole string and ends in an accept state.
Page 53	

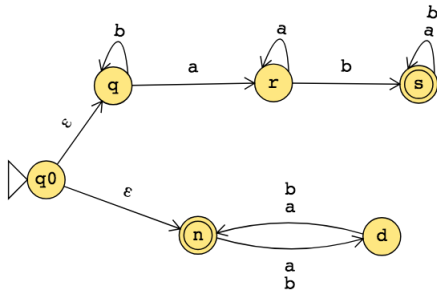
The formal definition of the NFA over  $\{0, 1\}$  given by this state diagram is:



The language over  $\{0, 1\}$  recognized by this NFA is:

Change the transition function to get a different NFA which accepts the empty string.

The state diagram of an NFA over  $\{a, b\}$  is below. The formal definition of this NFA is:



The language recognized by this NFA is:

## Week1 friday

**Review:** Determine whether each statement below about regular expressions over the alphabet  $\{a, b, c\}$  is true or false:

True or False:  $a \in L((a \cup b) \cup c)$

True or False:  $ab \in L((a \cup b)^*)$

True or False:  $ba \in L(a^*b^*)$

True or False:  $\varepsilon \in L(a \cup b \cup c)$

True or False:  $\varepsilon \in L((a \cup b)^*)$

True or False:  $\varepsilon \in L(a^*b^*)$

**From the pre-class reading, pages 34-36:** A deterministic finite automaton (DFA) is specified by  $M = (Q, \Sigma, \delta, q_0, F)$ . This 5-tuple is called the **formal definition** of the DFA. The DFA can also be represented by its state diagram: with nodes for the state, labelled edges specifying the transition function, and decorations on nodes denoting the start and accept states.

Finite set of states  $Q$  can be labelled by any collection of distinct names. Often we use default state labels  $q_0, q_1, \dots$

The alphabet  $\Sigma$  determines the possible inputs to the automaton. Each input to the automaton is a string over  $\Sigma$ , and the automaton “processes” the input one symbol (or character) at a time.

The transition function  $\delta$  gives the next state of the DFA based on the current state of the machine and on the next input symbol.

The start state  $q_0$  is an element of  $Q$ . Each computation of the machine starts at the start state.

The accept (final) states  $F$  form a subset of the states of the DFA,  $F \subseteq Q$ . These states are used to flag if the machine accepts or rejects an input string.

The computation of a machine on an input string is a sequence of states in the machine, starting with the start state, determined by transitions of the machine as it reads successive input symbols.

The DFA  $M$  accepts the given input string exactly when the computation of  $M$  on the input string ends in an accept state.  $M$  rejects the given input string exactly when the computation of  $M$  on the input string ends in a nonaccept state, that is, a state that is not in  $F$ .

The language of  $M$ ,  $L(M)$ , is defined as the set of all strings that are each accepted by the machine  $M$ . Each string that is rejected by  $M$  is not in  $L(M)$ . The language of  $M$  is also called the language recognized by  $M$ .

What is **finite** about all deterministic finite automata? (Select all that apply)

- ☐ The size of the machine (number of states, number of arrows)
- ☐ The number of strings that are accepted by the machine
- ☐ The length of each computation of the machine



The formal definition of this DFA is

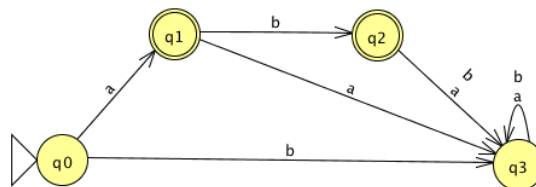
Classify each string  $a, aa, ab, ba, bb, \varepsilon$  as accepted by the DFA or rejected by the DFA.

*Why are these the only two options?*

The language recognized by this DFA is



The language recognized by this DFA is



The language recognized by this DFA is