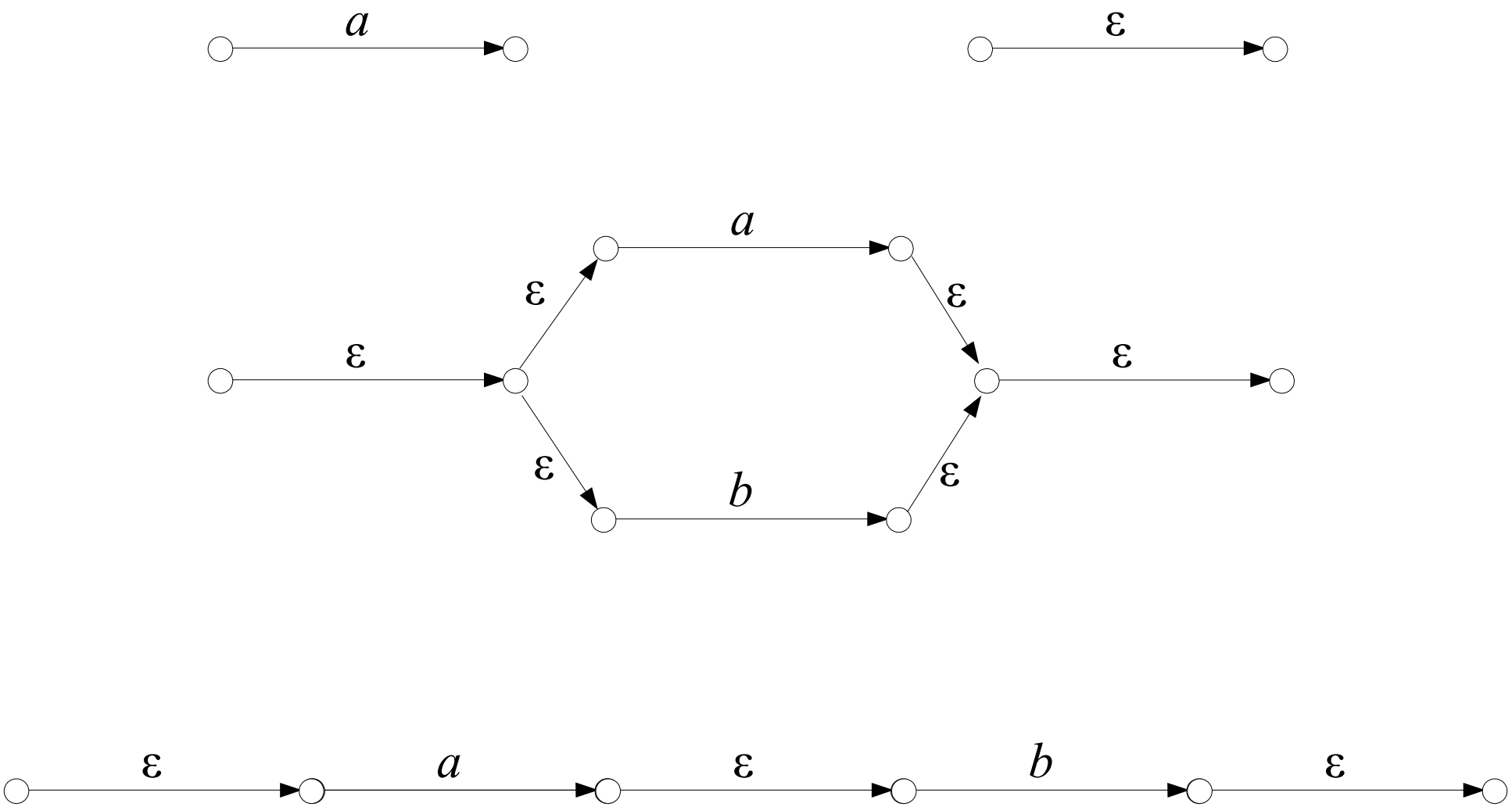
# Regular Expressions

Simon Law
23 January 2003

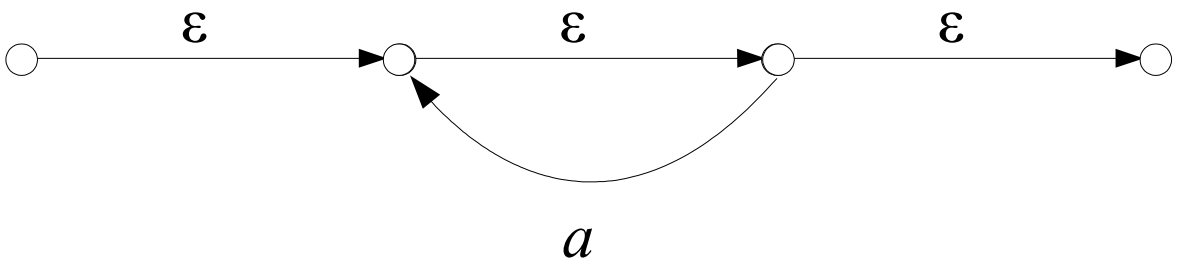
# Non-deterministic Finite Automata



# Algebra of regular sets

- Stephen Kleene discovered:
  - Alternation is commutative
  - Concatenation is non-commutative
  - Alternation is associative
  - Concatenation is associative
  - Distributive property holds

# Kleene Star



# Syntax

**a**

**a \| b**

**ab**

**a***

# Irregularities

`.`

`^`

`$`

`\w`

# Character classes

```
[ ]

[^ ]

[:alnum:] [:alpha:] [:blank:]
[:cntrl:] [:digit:] [:graph:]
[:lower:] [:print:] [:punct:]
[:space:] [:upper:] [:xdigit:]
```

# Character equivalents

**[=a=]**

**[[=a=]]**

matches

a á å ä

# Grouping

```
\( \)

\1 \2 \3

&
```

# Quantifiers

`+`

`?`

`{ }`

# Greedy

- Regular expressions are greedy
  - Match all **<br>** and change to **<br/>**

# Negation

- Negation is tricky
- **`x`**, so long as not followed by a slash
- a slash, so long as not preceded by **`x`**

Questions?