What voting gives us

- You might get system-wide agreement or you might get stuck
- Can’t vote directly on consensus question (i.e., log entry)
- What can we vote on without jeopardizing liveness?
  1. Statements that never get stuck (irrefutable), and
  2. Statements whose hold on consensus question can be broken if stuck (neutralizable)

Paxos [Lamport]

- A ballot is a pair \((n, x)\)
  - \(n\) – a counter to ensure arbitrarily many ballots exist
  - \(x\) – a candidate output value for the consensus protocol
- Conceptually vote to commit and abort ballots
  - If a quorum votes to commit \((n, x)\) for any \(n\), it is safe to output \(x\)
- Invariant: all committed and stuck ballots must have same \(x\)
- To preserve: can’t vote to commit a ballot before preparing it
  - Prepare \((n, x)\) by aborting all \((n’, x’)\) with \(n’ \leq n\) and \(x’ \neq x\).
  - PREPARED message votes to abort all lower ballots not containing \(x\)
    (or all lower ballots period if previous is NULL)
- If ballot \((n, x)\) stuck, neutralize by restarting with \((n + 1, x)\)
  - Can prepare \((n + 1, x)\) even if \((n, x)\) is stuck

Paxos example

Initial states:

- Any combination of bivalent and stuck ballots are possible
- Candidate values: 1 to 8
- Counter values: 1 to 4

1. Initially, all ballots are bivalent
   1. Agree that \((1, g)\) is prepared and vote to commit it
   2. Lose vote on \((1, g)\); agree \((2, f)\) prepared and vote to commit it
   3. \(2, f\) is stuck, so agree \((3, f)\) prepared and vote to commit it
   4. See \(T\) votes to commit \((3, f)\) (commit-valent) and externalize \(f\)
      - At this point nobody cares about \((2, f)\) — neutralized
   5. Node failure makes \((3, f)\) stuck, prepare and commit \((4, f)\)

2. Initially, all ballots are bivalent
   1. Agree that \((1, g)\) is prepared and vote to commit it
   2. Lose vote on \((1, g)\); agree \((2, f)\) prepared and vote to commit it
   3. \(2, f\) is stuck, so agree \((3, f)\) prepared and vote to commit it
   4. See \(T\) votes to commit \((3, f)\) (commit-valent) and externalize \(f\)
      - At this point nobody cares about \((2, f)\) — neutralized
   5. Node failure makes \((3, f)\) stuck, prepare and commit \((4, f)\)
### Paxos example

<table>
<thead>
<tr>
<th>candidate values</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

0. Initially, all ballots are bivalent
1. Agree that \( (1, g) \) is prepared and vote to commit it
2. Lose vote on \( (1, g) \); agree \( (2, f) \) prepared and vote to commit it
3. \( (2, f) \) is stuck, so agree \( (3, f) \) prepared and vote to commit it
4. See \( T \) votes to commit \( (3, f) \) (commit-valent) and externalize \( f \)
   - At this point nobody cares about \( (2, f) \)—neutralized
5. Node failure makes \( (3, f) \) stuck, prepare and commit \( (4, f) \)

### Viewstamped replication [Oki]

**view 1:** \( \text{op}_1 \text{ } \text{op}_2 \text{ } \text{op}_3 \text{ } \text{op}_4 \text{ } ? \)

- Instead of voting on \( \text{op}_1 \), . . . directly, vote on \( \langle \text{view } 1, \text{op}_1 \rangle \), . . .
  - Each \( \langle \text{view}, \text{op} \rangle \) selected by a single leader for view, so irrefutable
  - E.g., chose leader by round-robin using view\# mod \( N \)

- What if votes on \( \text{op}_3 \) and \( \text{op}_3 \) are stuck (e.g., leader fails)?
  - Neutralize by agreeing view 1 had only 3 meaningful operations
  - Vote to form view 2 that immediately follows \( \langle \text{view } 1, \text{op}_3 \rangle \)

- Failed to form view 2 (e.g., because a node wants \( \langle \text{view } 1, \text{op}_3 \rangle \))?  
  - Just go on to form view 3 after \( \langle \text{view } 1, \text{op}_3 \rangle \)

**view 2:** \( \text{op}_4 \text{ } \text{op}_5 \text{ } \text{op}_6 \text{ } \ldots \)

**view 3:** \( \text{op}_5 \text{ } \text{op}_6 \text{ } \ldots \)

- Instead of voting on \( \text{op}_1 \), . . . directly, vote on \( \langle \text{view } 1, \text{op}_1 \rangle \), . . .
  - Each \( \langle \text{view}, \text{op} \rangle \) selected by a single leader for view, so irrefutable
  - E.g., chose leader by round-robin using view\# mod \( N \)

- What if votes on \( \text{op}_4 \) and \( \text{op}_5 \) are stuck (e.g., leader fails)?
  - Neutralize by agreeing view 1 had only 3 meaningful operations
  - Vote to form view 2 that immediately follows \( \langle \text{view } 1, \text{op}_3 \rangle \)

- Failed to form view 2 (e.g., because a node wants \( \langle \text{view } 1, \text{op}_3 \rangle \))?  
  - Just go on to form view 3 after \( \langle \text{view } 1, \text{op}_3 \rangle \)