

Git: Advanced Topics

Computer Science and Engineering ■ College of Engineering ■ The Ohio State University

Lecture 4

Basic Workflow: Overview

1. Configure git locally (everyone, once)
2. Create central repo (1 person)
3. Create local repo (everyone)
4. Local development (everyone):
 - Commit locally
 - Fetch/merge as appropriate
 - Push to share

Step 1: Configure Git Locally

- Each team member, on their own machine
 - Required: Set identity for authoring commits

```
$ git config --global user.name "Brutus Buckeye"
```

```
$ git config --global user.email bb@osu.edu
```
 - Recommend: set default initial branch name (2.28+)

```
$ git config --global init.defaultBranch main
```
 - Tips
 - Add email to GitHub account (Settings > Email)
 - Alternative: use GitHub-generated fake address:
 - Settings > Email > Keep my address private
 - Find `ID+USERNAME@users.noreply.github.com`
 - Add your public SSH key to your GitHub account

Step 2: Initialize Central Rep

- ❑ One person, once per project
- ❑ Hosting services (GitHub, GitLab, BitBucket...) use a web interface for this step
 - In 3901, this will be done for you
- ❑ No magic: any location ok, so long as the group can access (*e.g.* coelinux):
 - Create central repository in group's project directory (/project/c3901aa03)

```
$ cd /project/c3901aa03
```
 - \$ **mkdir** proj1 *# an ordinary directory*
 - Initialize directory as a *bare* git repository, with group permissions
 - \$ **git init** --bare --shared proj1

Step 3: Create Local Repository

- Each team member, once, on their machine
 - Create local repo by *cloning* the central one

```
$ git clone git@github.com:bb/proj1.git
```
 - Copies entire repo, including store, and sets a remote called "origin"

```
$ cd proj1
```

```
proj1$ git remote -v # display info
```

```
origin git@github.com:bb/proj1.git (fetch)
```

```
origin git@github.com:bb/proj1.git (push)
```
- Different ways to clone
 - SSH
 - HTTPS with Git Credential Manager

Step 4: Local Development

□ Each team member repeats:

- Edit and commit (to local repository) often

```
$ git status/add/rm/commit
```

- Pull others' work when you can benefit

```
$ git fetch origin # bring in changes
```

```
$ git log/checkout # examine new work
```

```
$ git merge, commit # merge work
```

- Push to central repository when confident

```
$ git push origin main # share
```

Demo: Local Development

□ Playground:

- rcmarques3.github.io/visualizing-git/#upstream-changes

□ Try:

```
git commit
```

```
git fetch origin # see origin/feature
```

```
git merge origin/feature # see feature
```

```
git push origin feature # see remote
```

Your Turn: Playing with Git

- ❑ Navigate to class org on GH and find the repo called *first-commits*
- ❑ Clone the repo to your machine
- ❑ Do some development!
 - Edit
 - Inspect the store's DAG

```
$ git log --graph --oneline --all
```
 - Commit, fetch, merge, push...
 - Rinse, repeat

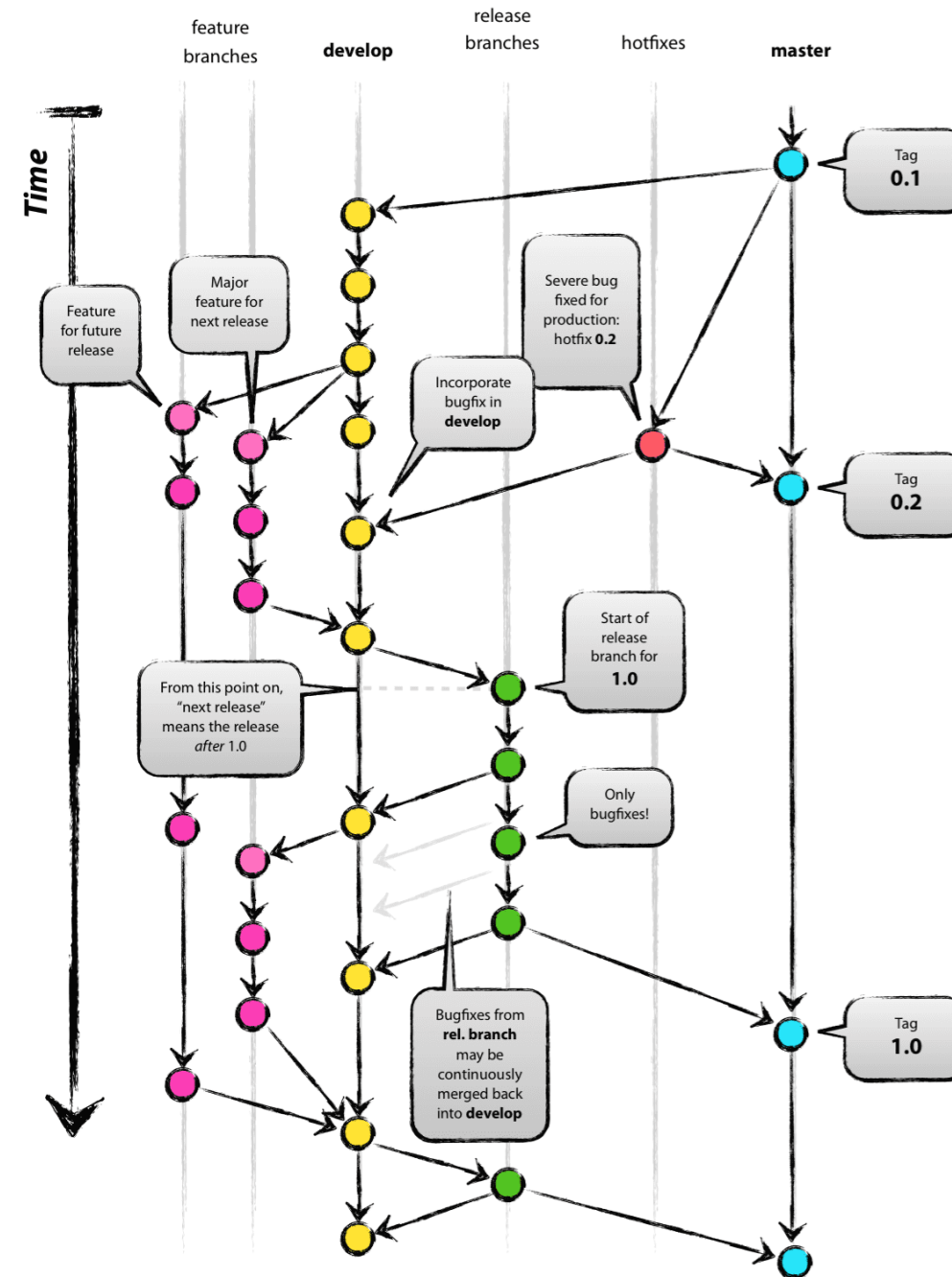
Professional Git

- ❑ Commit/branch conventions
- ❑ Deciding what goes in, and what stays out of the store
 - Share all the things that should be shared
 - Only share things that should be shared
- ❑ Normalizing contents of the store
 - Windows vs linux line endings

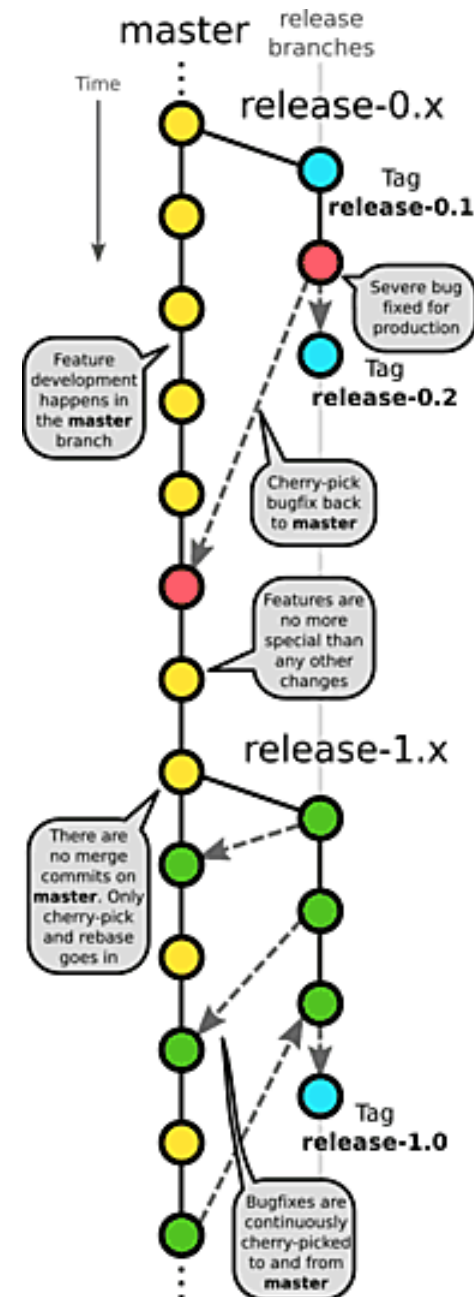
Commit/Branch Conventions

- ❑ Team strategy for managing the DAG (ie the store)
- ❑ Examples:
 - “Main is always deployable”
 - ❑ All work is done on other branches, merged with main only when result is executable
 - “Feature branches”, “developer branches”
 - ❑ Each feature developed on its own branch vs. each developer works on their own branch
 - “Favor rebase over merge”
 - ❑ Always append to latest origin/branch

Example: Branch-Based Dev



Example: Trunk-Based Dev



What Goes Into Central Repo?

- ❑ Avoid private information
 - Passwords, secret tokens
 - Better: Use *environment variables* instead
- ❑ Avoid developer-specific environment settings
 - Hard-coded file/directory paths from local machine
 - OK to include a sample config (each developer customizes but keeps their version out of store)
- ❑ Avoid living binaries (docx, pdf)
 - Meaningless diffs
- ❑ Avoid generated files
 - compiled files, the build
- ❑ Avoid IDE-specific files (.settings)
 - Some generic ones are OK so it is easier to get started by cloning, especially if the team uses the same IDE
- ❑ Agree on code formatting
 - Auto-format is good, but only if everyone uses the same format settings!
 - Spaces vs tabs, brace position, etc

Ignoring Files from Working Tree

- Use a .gitignore file in the root of project
 - Committed as part of the project
 - Consistent policy for everyone on team
- Examples: <https://github.com/github/gitignore>

```
# github:gitignore/Java.gitignore
# Compiled class file
*.class
```

```
# Log file
*.log
```

```
# Package Files #
*.jar
*.war
*.ear
*.zip
*.tar.gz
*.rar
```

Problem: End-of-line Confusion

- Differences between OS's in how a new line is encoded in a text file
 - Windows: 2 bytes, CR + LF ("`\r\n`", 0x0D 0x0A)
 - Unix/Mac: 1 byte, LF ("`\n`", 0x0A)
- Difference is hidden by most editors
 - An IDE might recognize either when opening a file, but convert all to `\r\n` when saving
 - Demo: hexdump (or VSCode hex editor)
- But difference matters to git when comparing files!
- Problem: OS differences within team
 - Changing 1 line causes every line to be modified
 - Flood of spurious changes masks the real edit

Solution: Normalization

- Convention: Store uses `\n` (ie linux)
 - Working tree uses OS's native eol
 - Convert when moving data between the two (e.g., commit, checkout)
- Note: Applies to text files only
 - A binary file, like a jpg, might contain `0x0D` and/or `0x0A`, but they should never be converted
- How does git know whether a file is text or binary?
 - Heuristics: auto-detect based on contents
 - Configuration: filename matches a pattern

Normalization With .gitattributes

- Use a .gitattributes file in root of project
 - Committed as part of the project
 - Consistent policy for everyone on team

- Example:

```
# Auto detect text files and perform LF normalization  
* text=auto
```

```
# These files are text, should be normalized (crlf=>lf)  
*.java      text  
*.md        text  
*.txt       text  
*.classpath text  
*.project   text
```

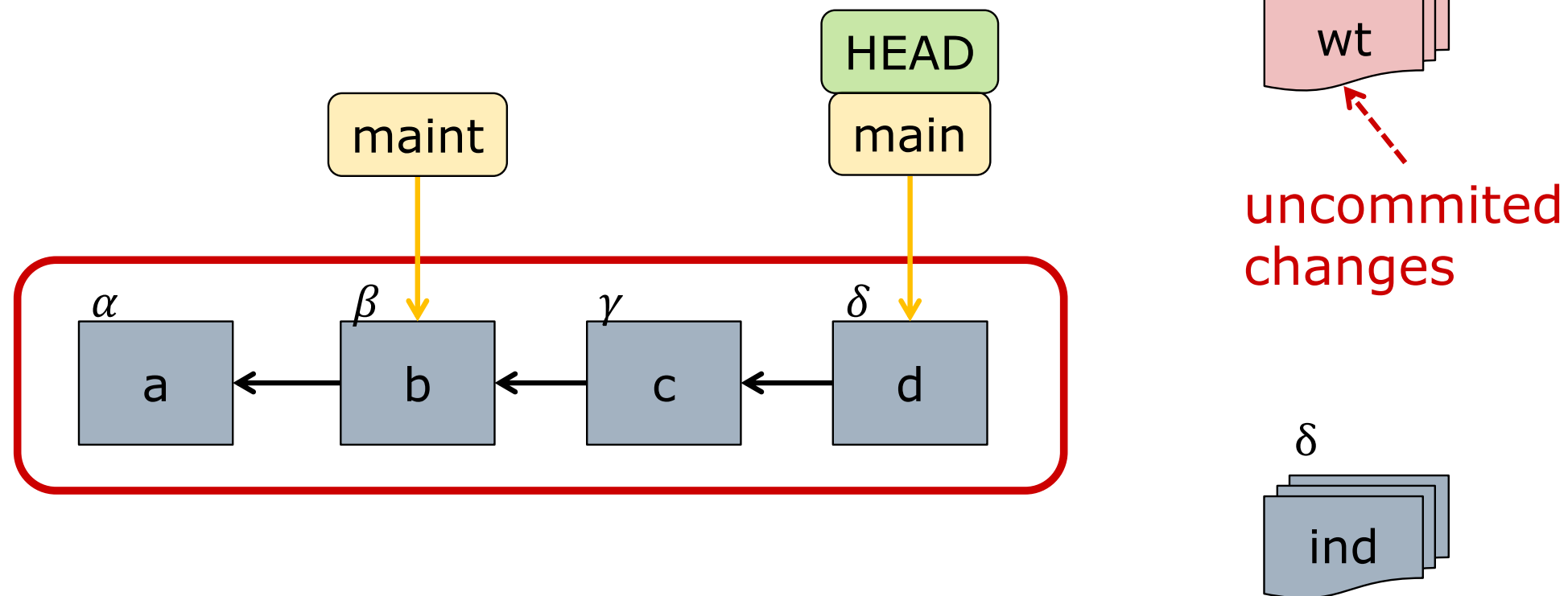
```
# These files are binary, should be left untouched  
*.class     binary  
*.jar       binary
```

Ninja Git: Advanced Moves

- Temporary storage
`stash`
- Undoing big and small mistakes in the working tree
`reset`, `checkout`
- Undoing mistakes in store
`amend`
- DAG surgery
`rebase`

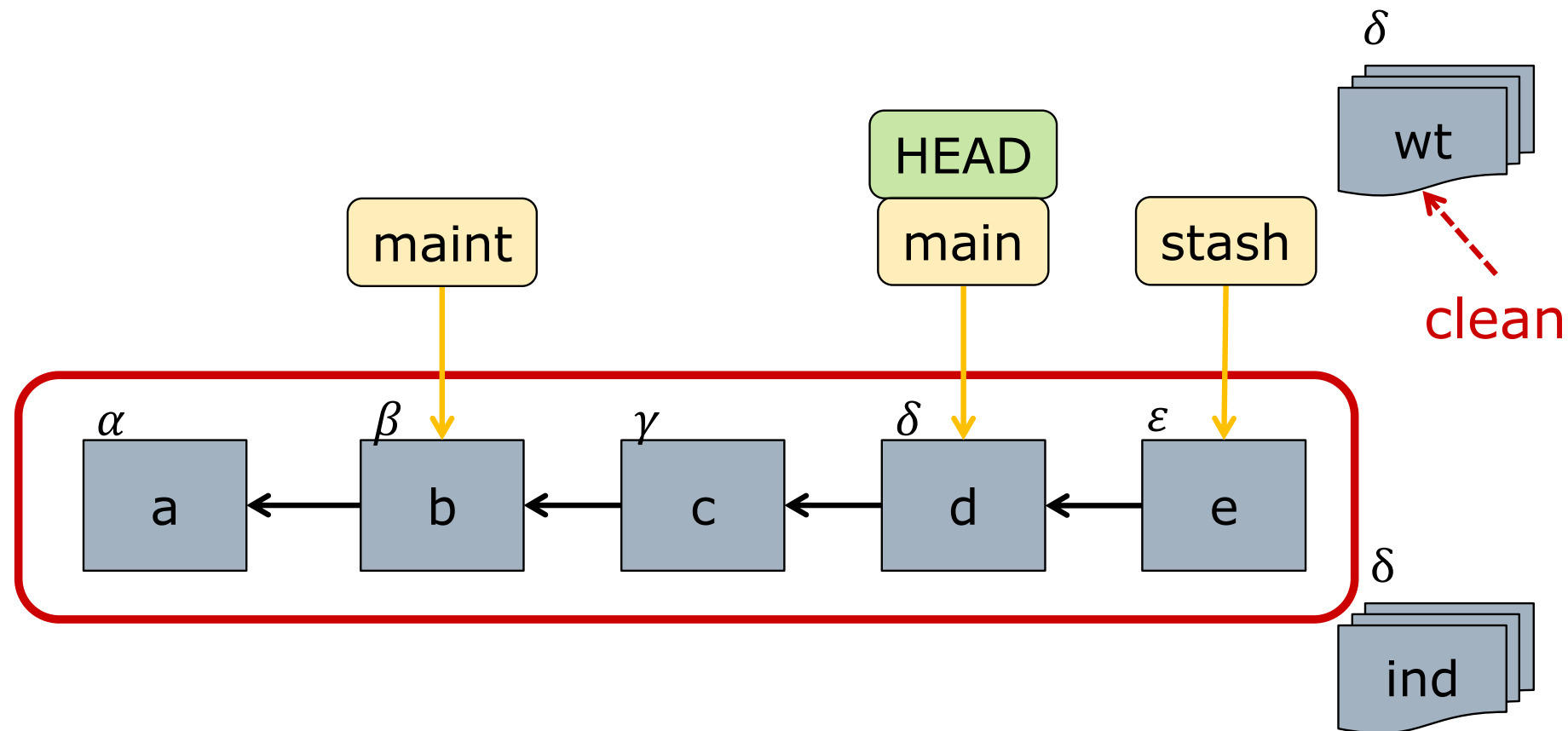
Advanced: Temporary Storage

- ❑ Say you have uncommitted work and want to look at a different branch
- ❑ Checkout won't work! (Recall rule: "only checkout when wt is clean")



Stash: Push Work Onto a Stack

```
$ git stash          # repo now clean  
$ git checkout ...etc... # feel free to poke around
```



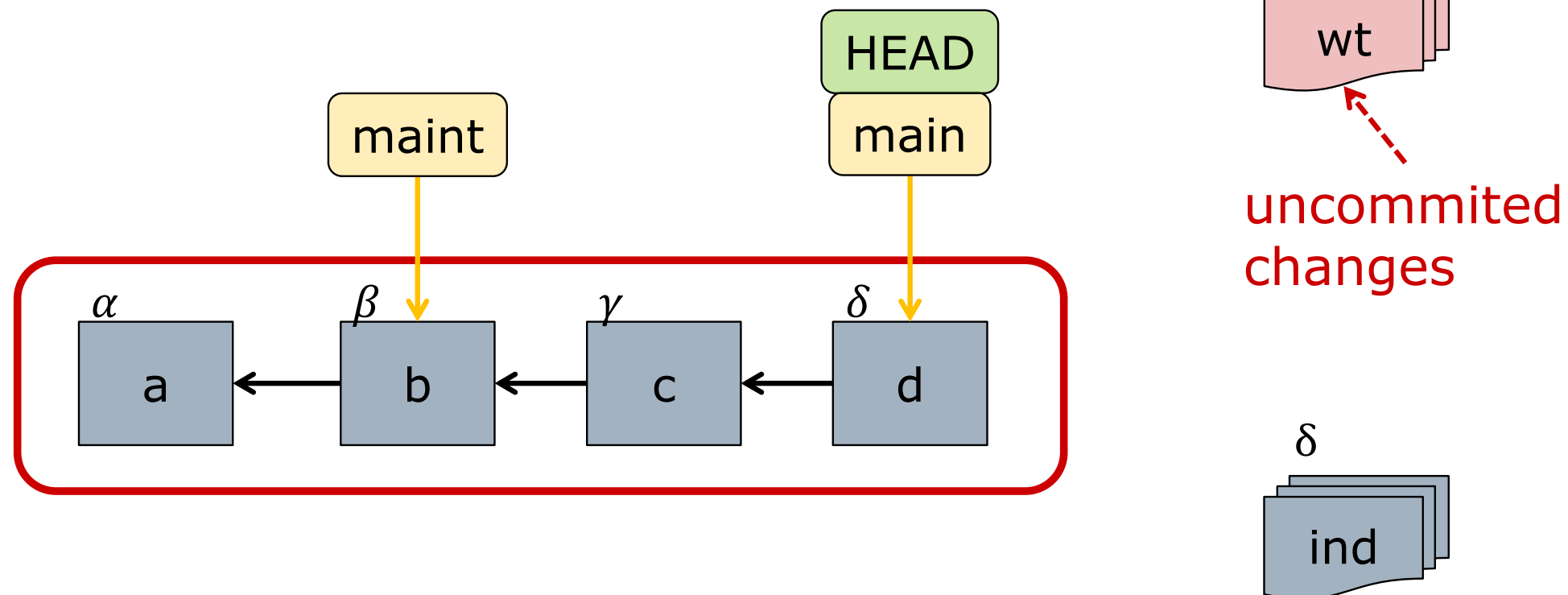
Stash: Pop Work Off the Stack

```
$ git stash pop    # restores state of wt/index
```

```
# one-liner above equivalent to:
```

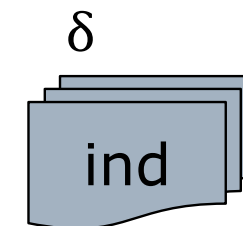
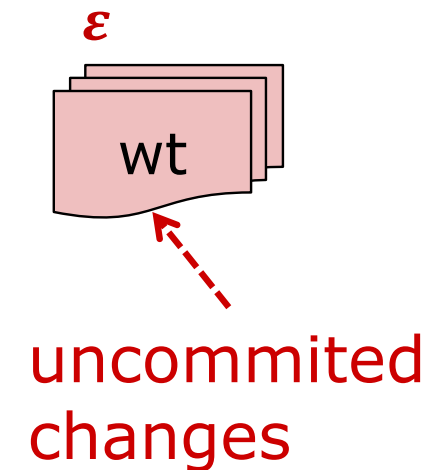
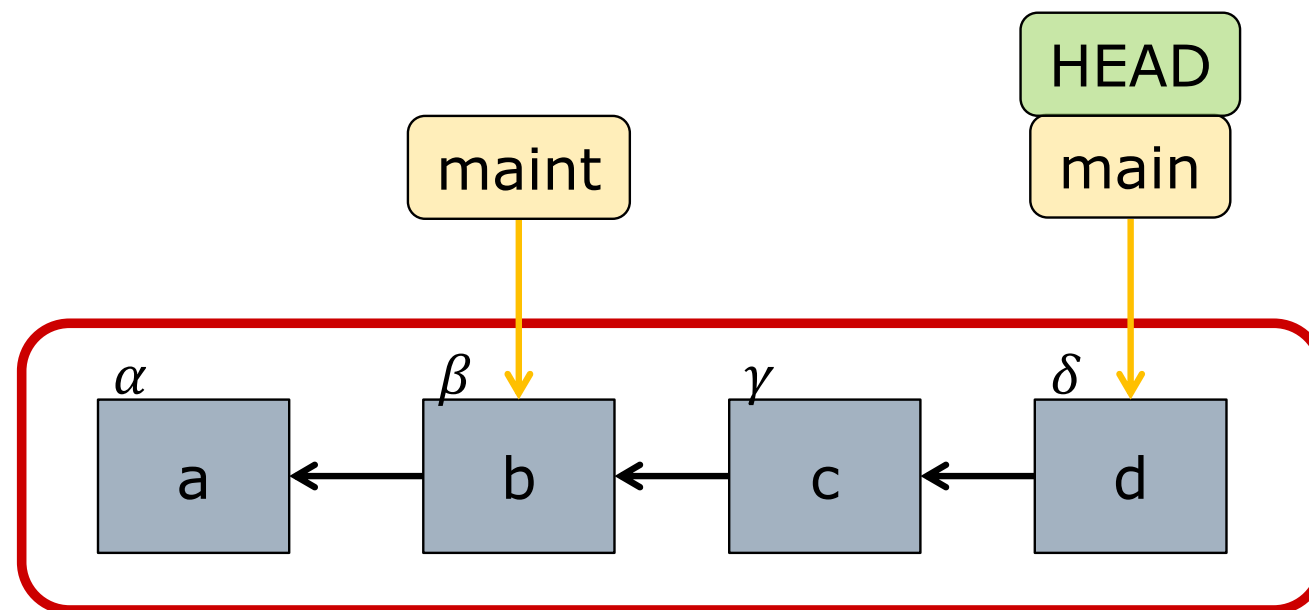
```
$ git stash apply # restore wt and index
```

```
$ git stash drop  # restore store
```



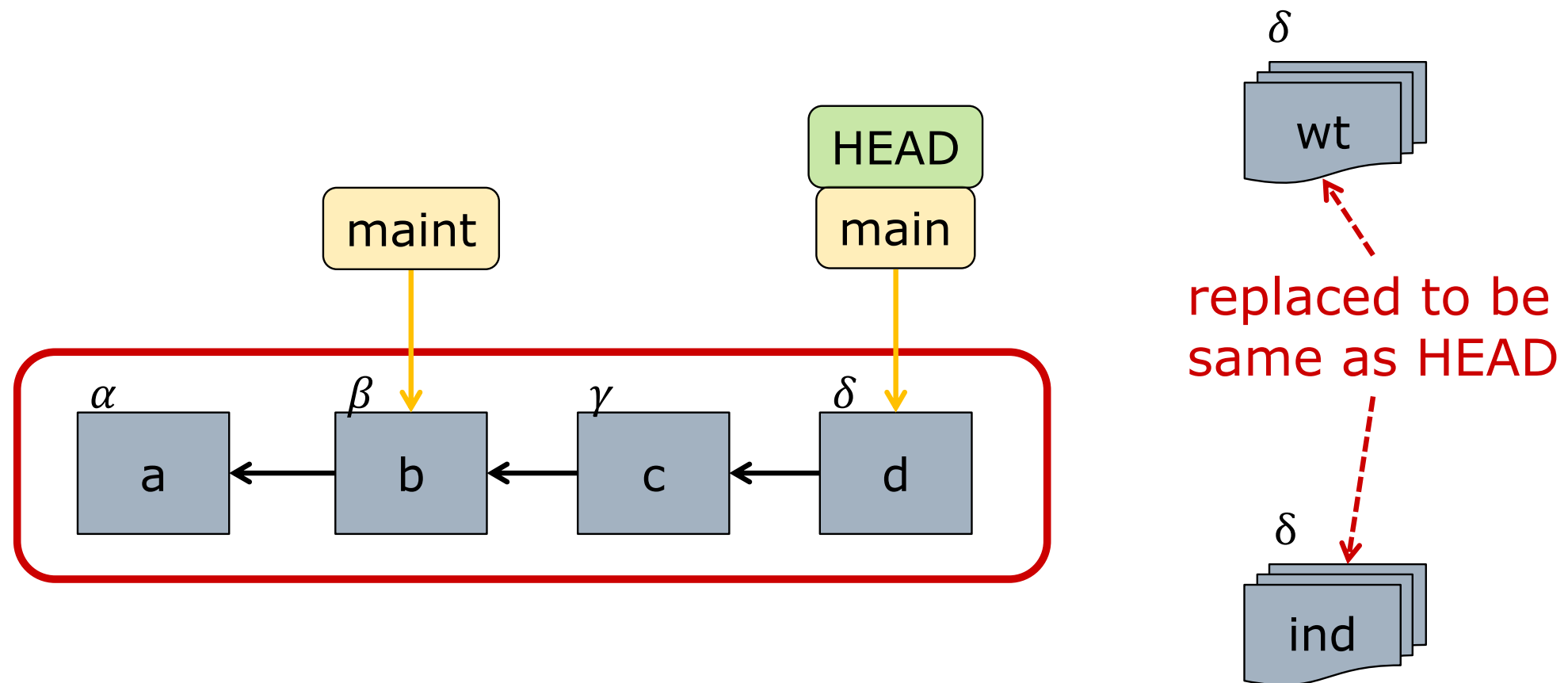
Advanced: Undoing Big Mistakes

- Say you want to throw away *all* your uncommitted work
 - ie “Roll back” to last committed state
- Checkout HEAD won't work!



Reset: Discarding Changes

```
$ git reset --hard      # updates wt to be HEAD
$ git clean --dry-run   # list untracked files
$ git clean --force     # remove untracked files
```

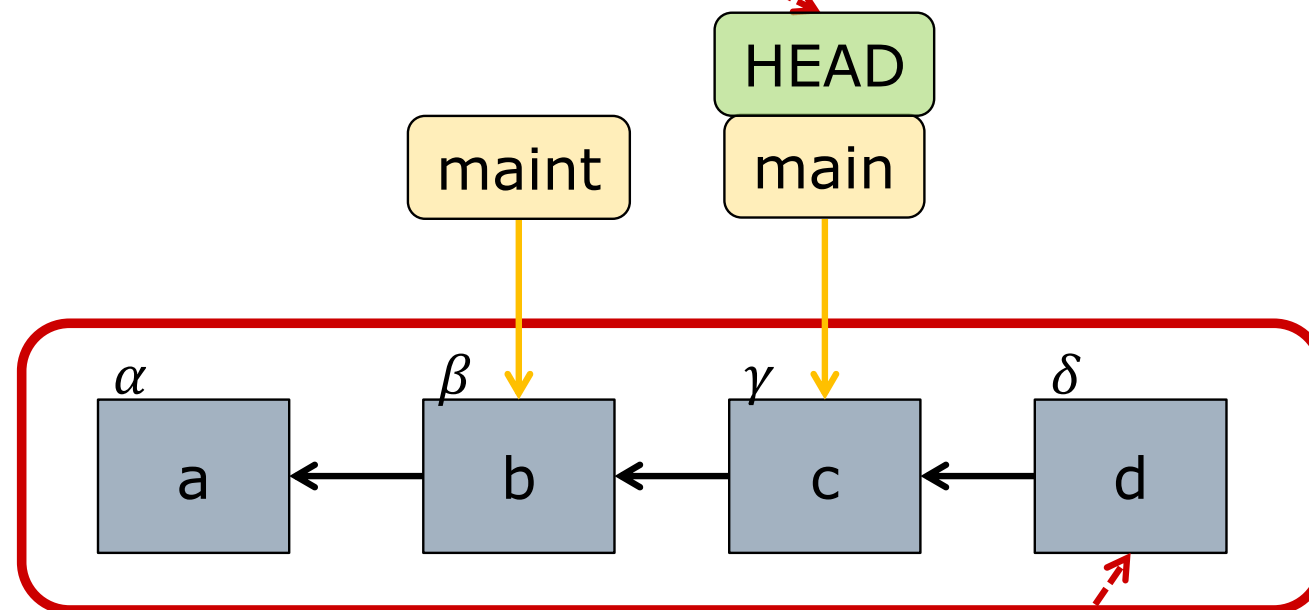


Reset: Discarding Commits

```
$ git reset --hard HEAD~1
```

no need to git clean, since wt was already clean

HEAD moved
(and attached branch)



now unreachable

γ

Diagram showing the working tree (wt) state. A stack of blue boxes represents the working tree, with the top box labeled 'wt'. A dashed red arrow points from the text 'replaced to be same as HEAD~1' to the 'wt' box.

replaced to be
same as
HEAD~1

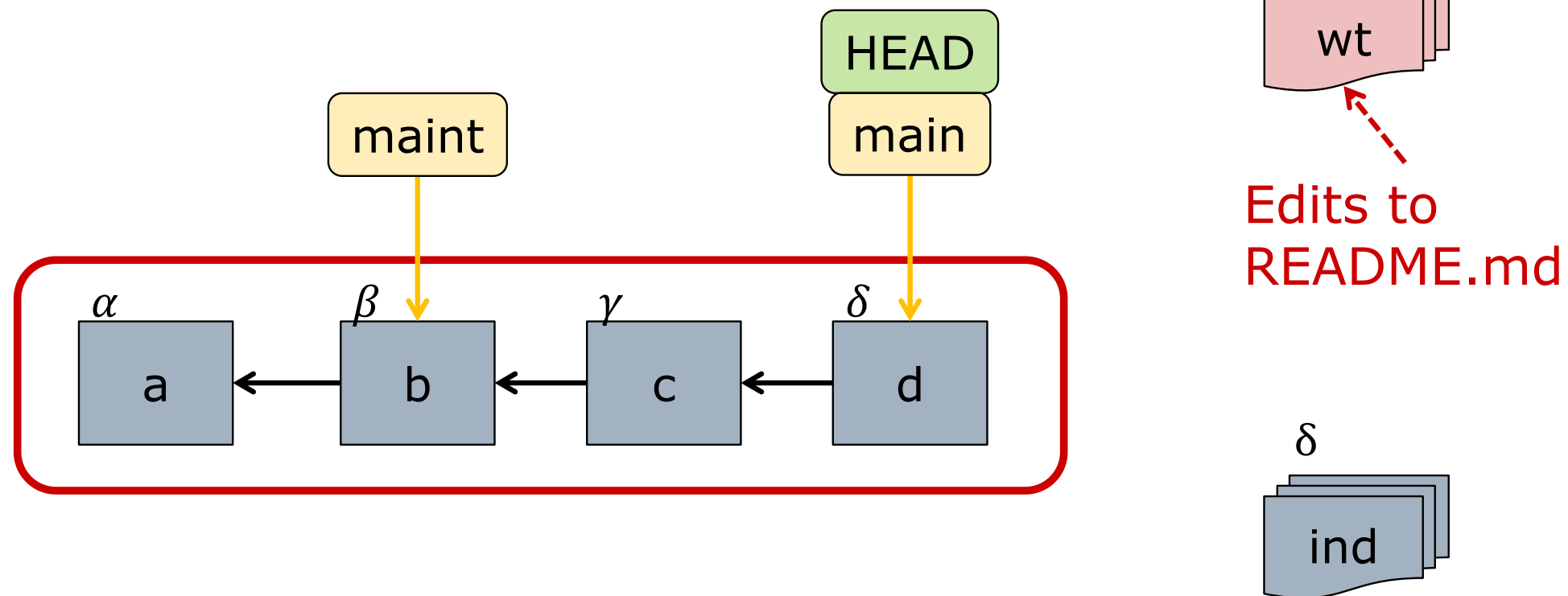
γ

Diagram showing the index (ind) state. A stack of blue boxes represents the index, with the top box labeled 'ind'. A dashed red arrow points from the text 'replaced to be same as HEAD~1' to the 'ind' box.

ind

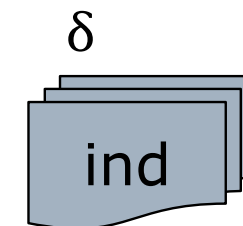
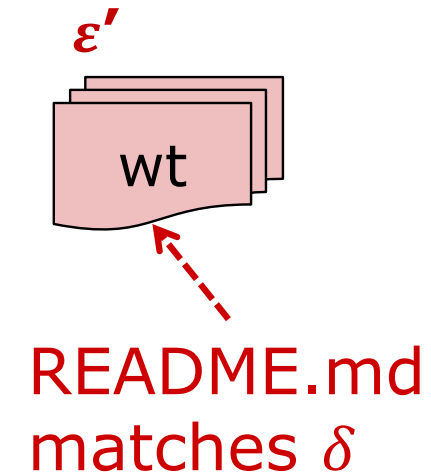
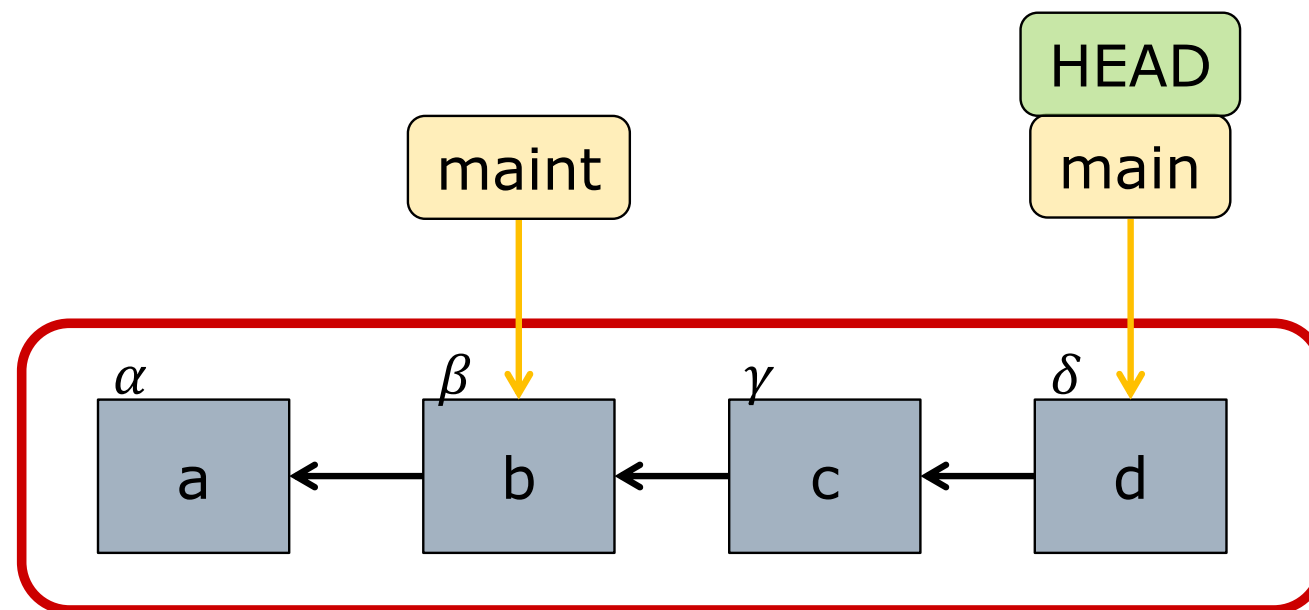
Advanced: Undo Small Mistakes

- Say you want to throw away *some of* your uncommitted work
 - Restore a file to last committed version



Advanced: Undo Small Mistakes (Checkout)

```
$ git checkout -- README.md
# -- means: rest is file/path (not branch)
# git checkout README.md ok, if not ambiguous
```



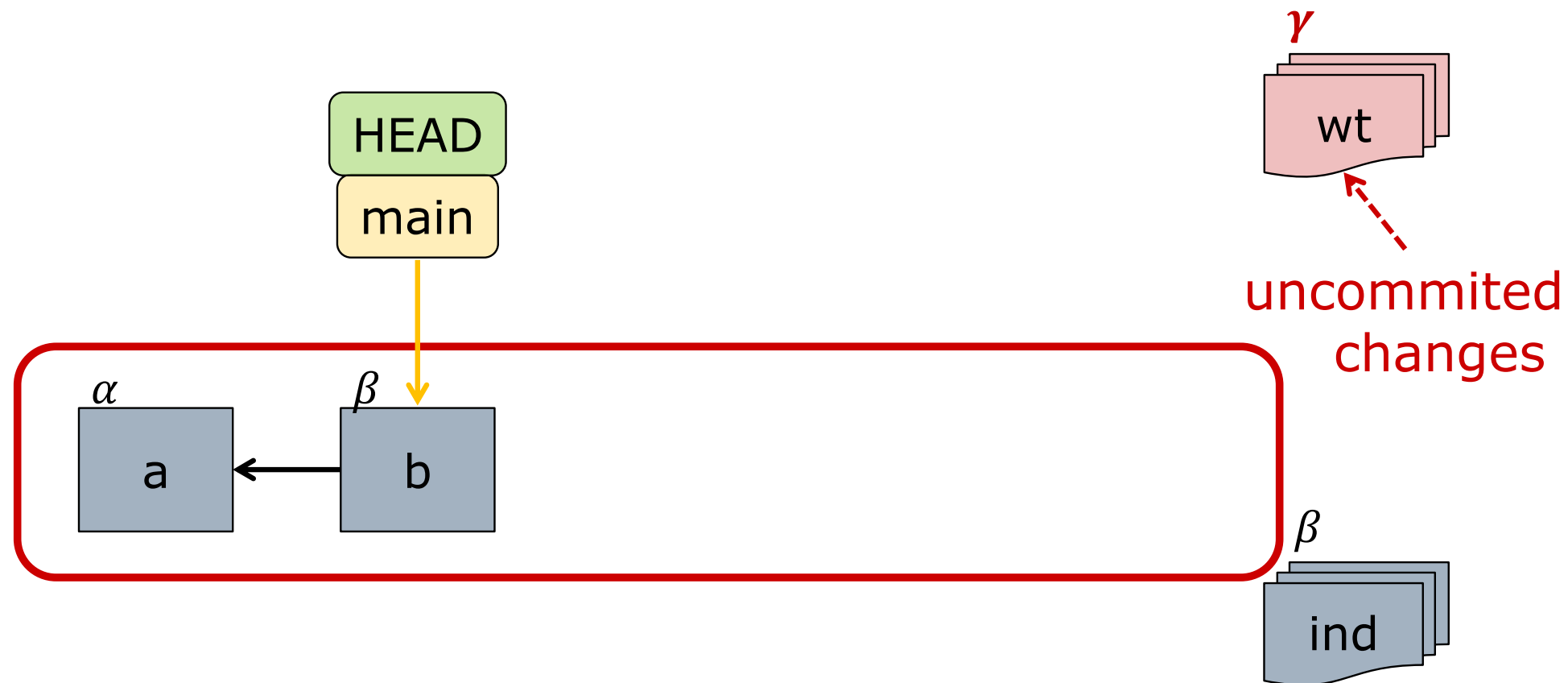
The Power to Change History

- Changing the store lets us:
 - Fix mistakes in recent commits
 - Clean up messy DAGs to make history look more linear

- Rule: Never change *shared* history
 - Once something has been pushed to a remote repo (*e.g.*, origin), do not change that part of the DAG
 - So: A *push* is really a *commitment*!

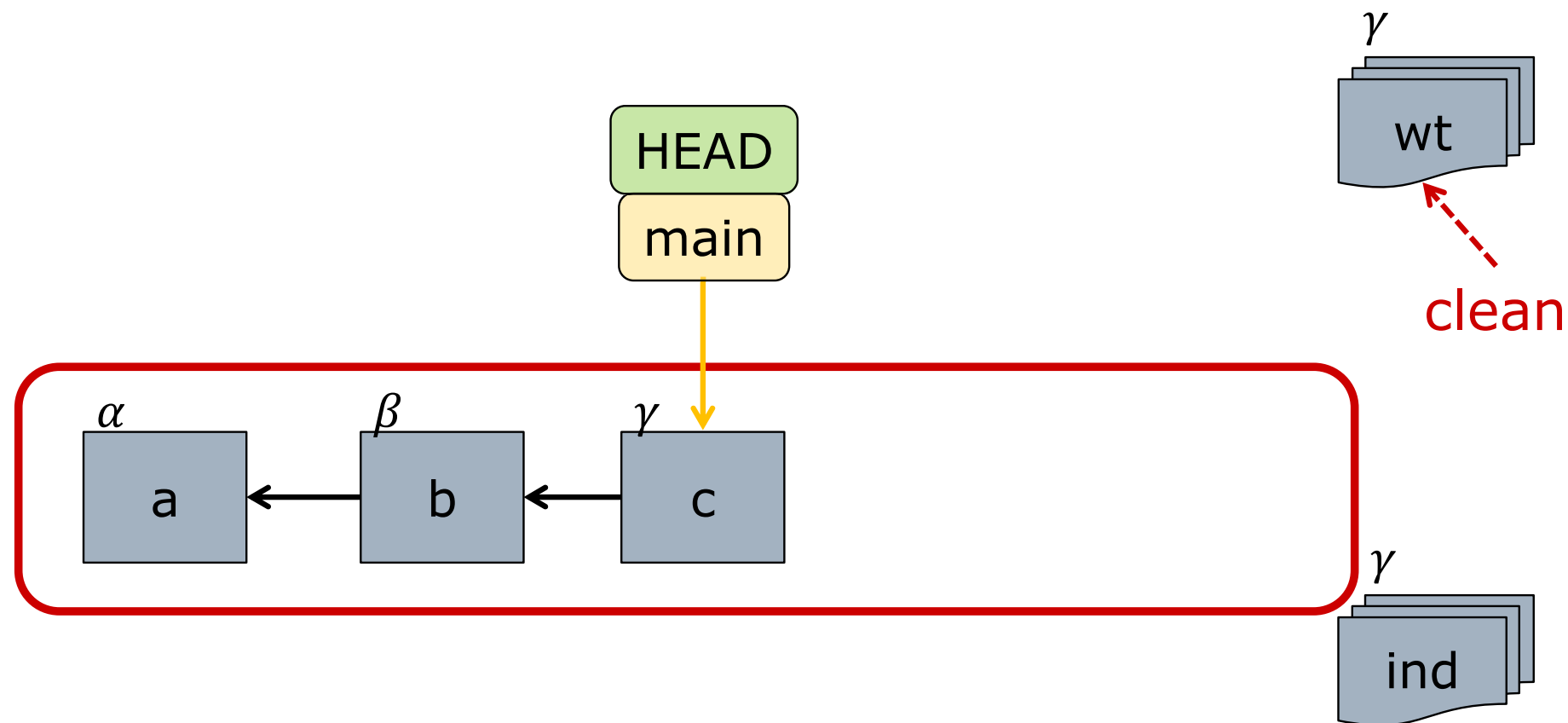
Problem 1: Wrong Commit

□ Problem 1: Wrong or incomplete commit



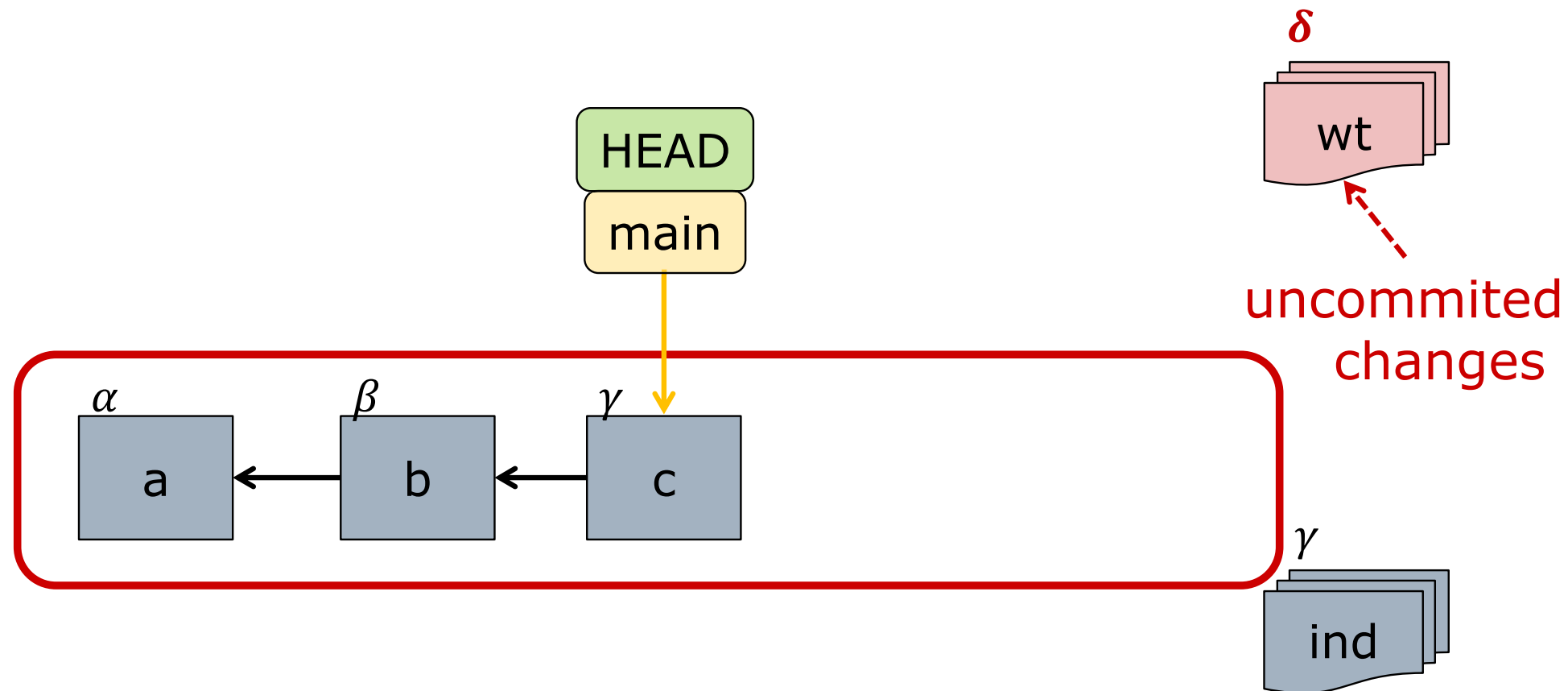
Problem 1: Wrong Commit (2)

□ Problem 1: Wrong or incomplete commit



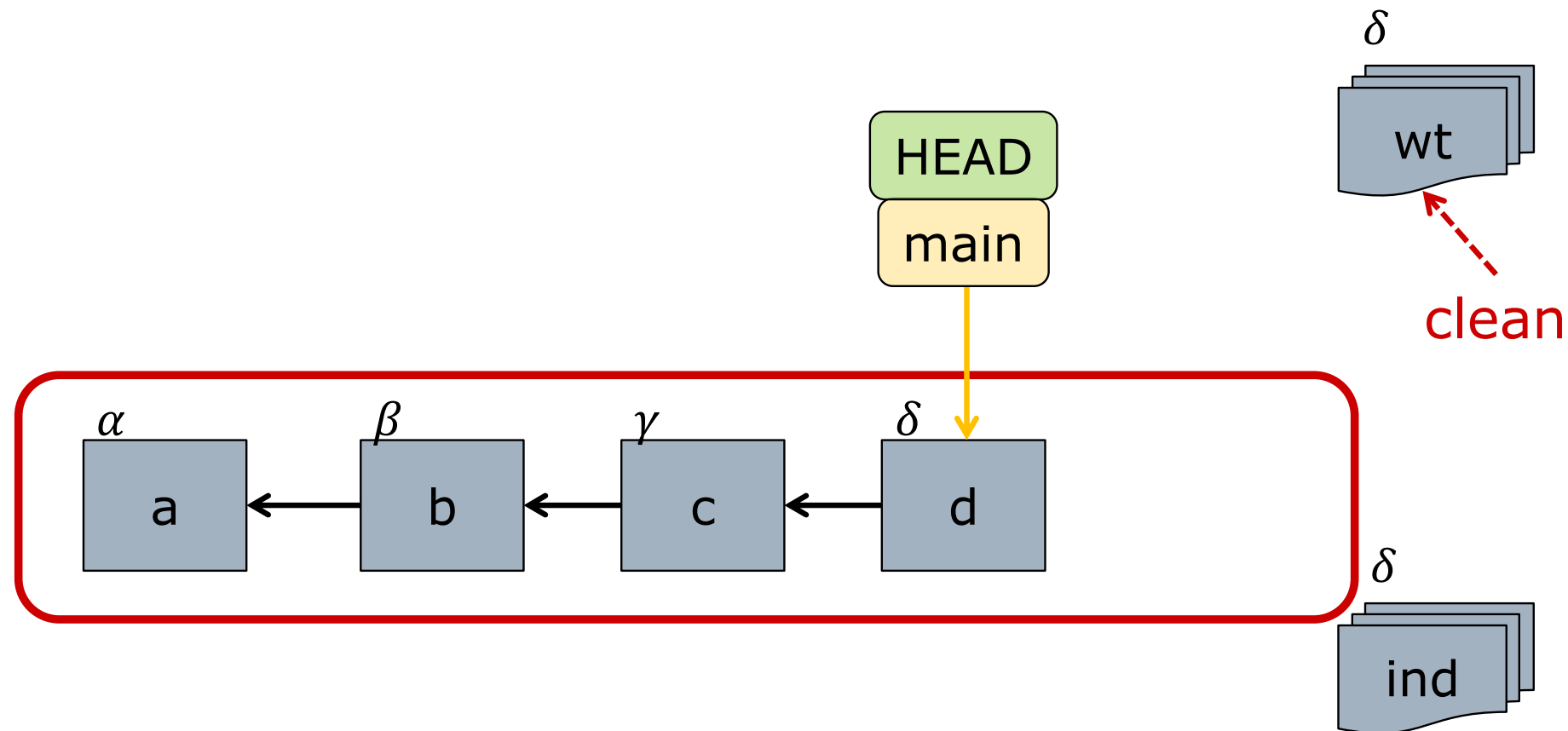
Problem 1: Wrong Commit (3)

- Problem 1: Wrong or incomplete commit
 - Oops! That wasn't quite right...



Problem 1: Wrong Commit (4)

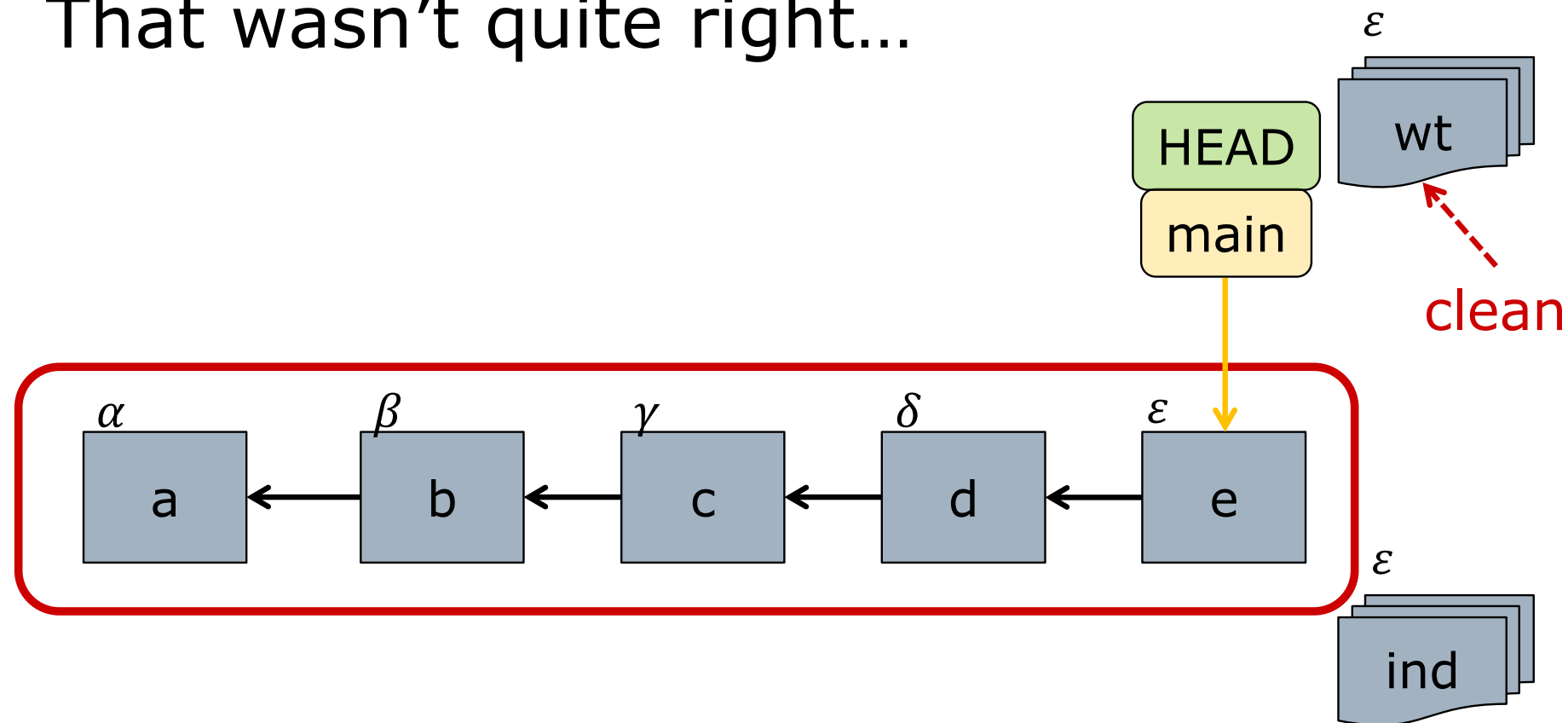
- ❑ Problem 1: Wrong or incomplete commit
 - Oops! That wasn't quite right...
 - Oops! That wasn't quite right...



Problem 1: Wrong Commit (5)

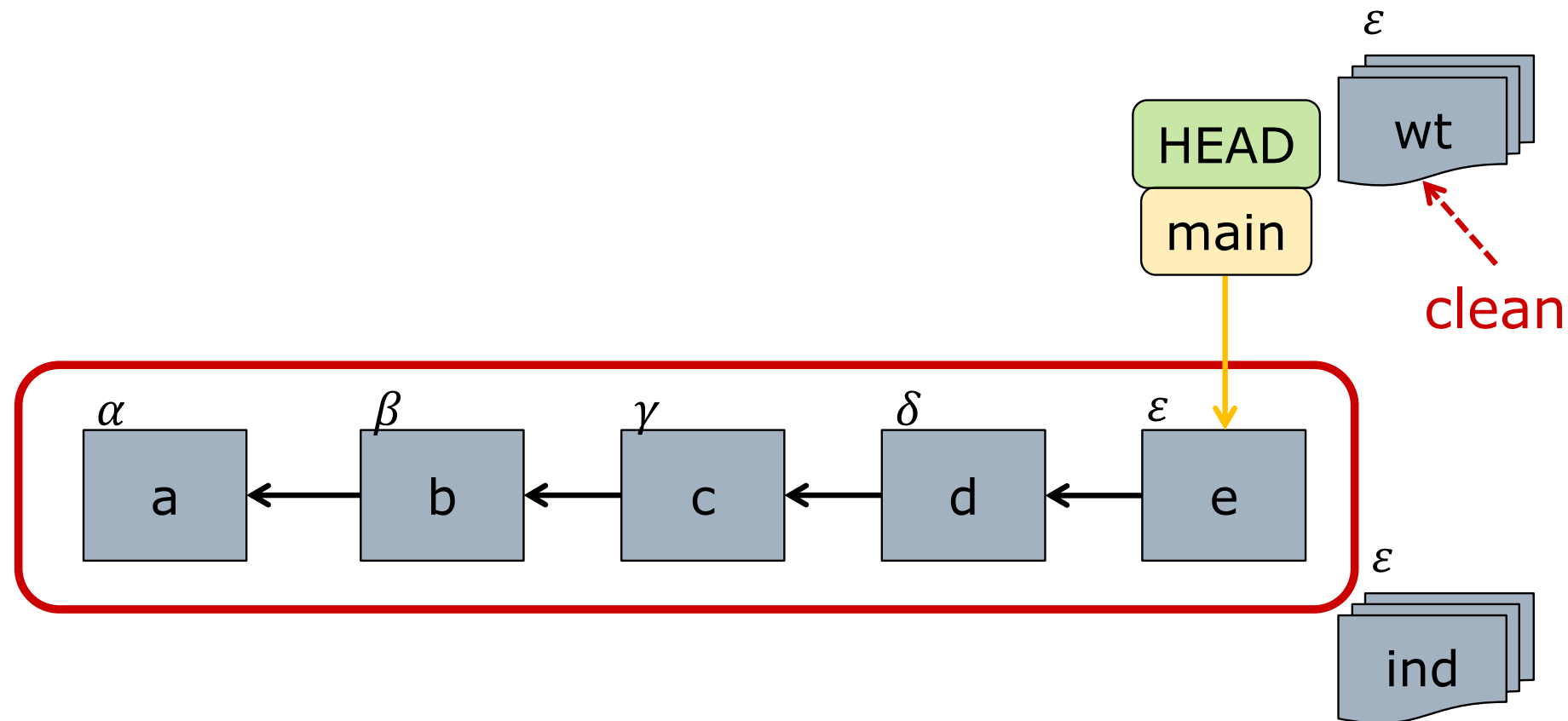
□ Problem 1: Wrong or incomplete commit

- Oops! That wasn't quite right...
- Oops! That wasn't quite right...
- Oops! That wasn't quite right...



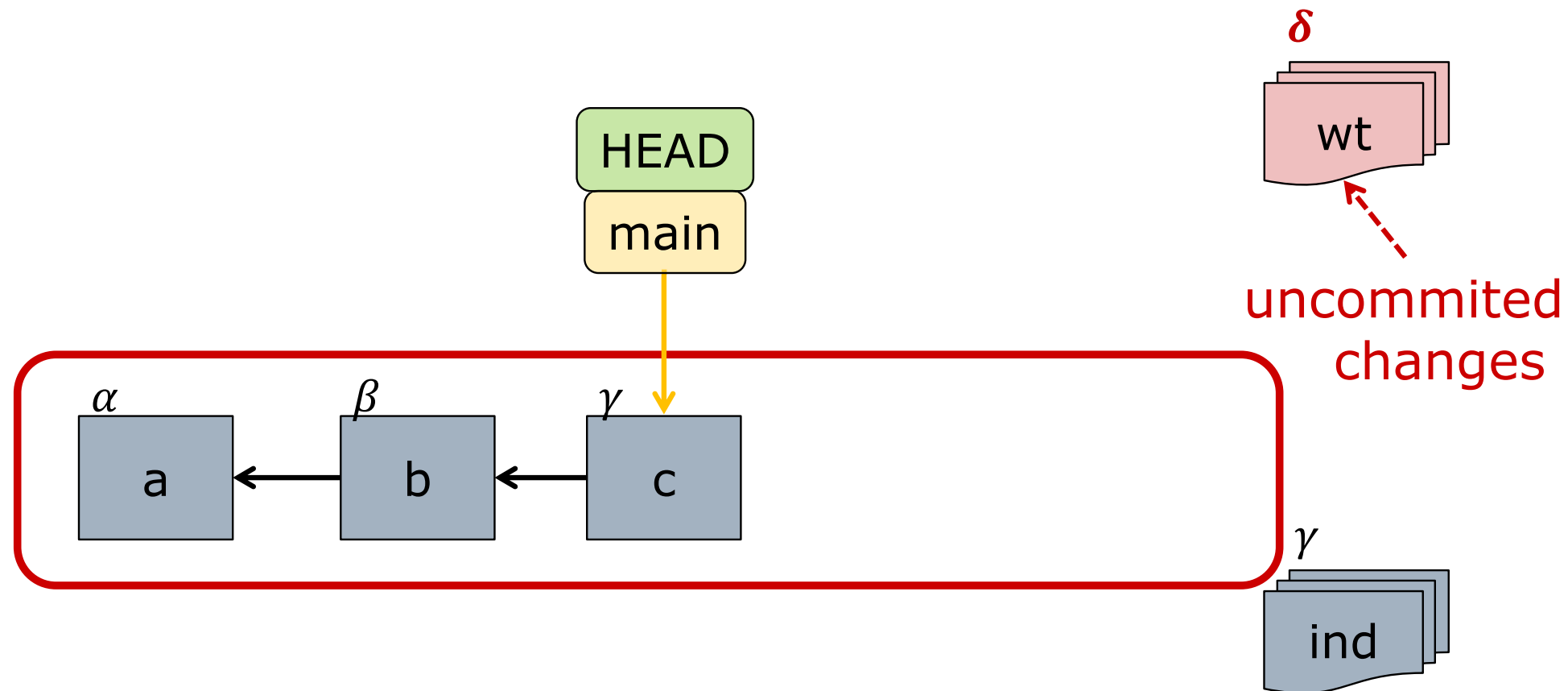
Problem 1: Wrong Commit (6)

- ❑ Problem 1: Wrong or incomplete commit
- ❑ Result: Lots of tiny “fix it”, “oops”, “retry” commits



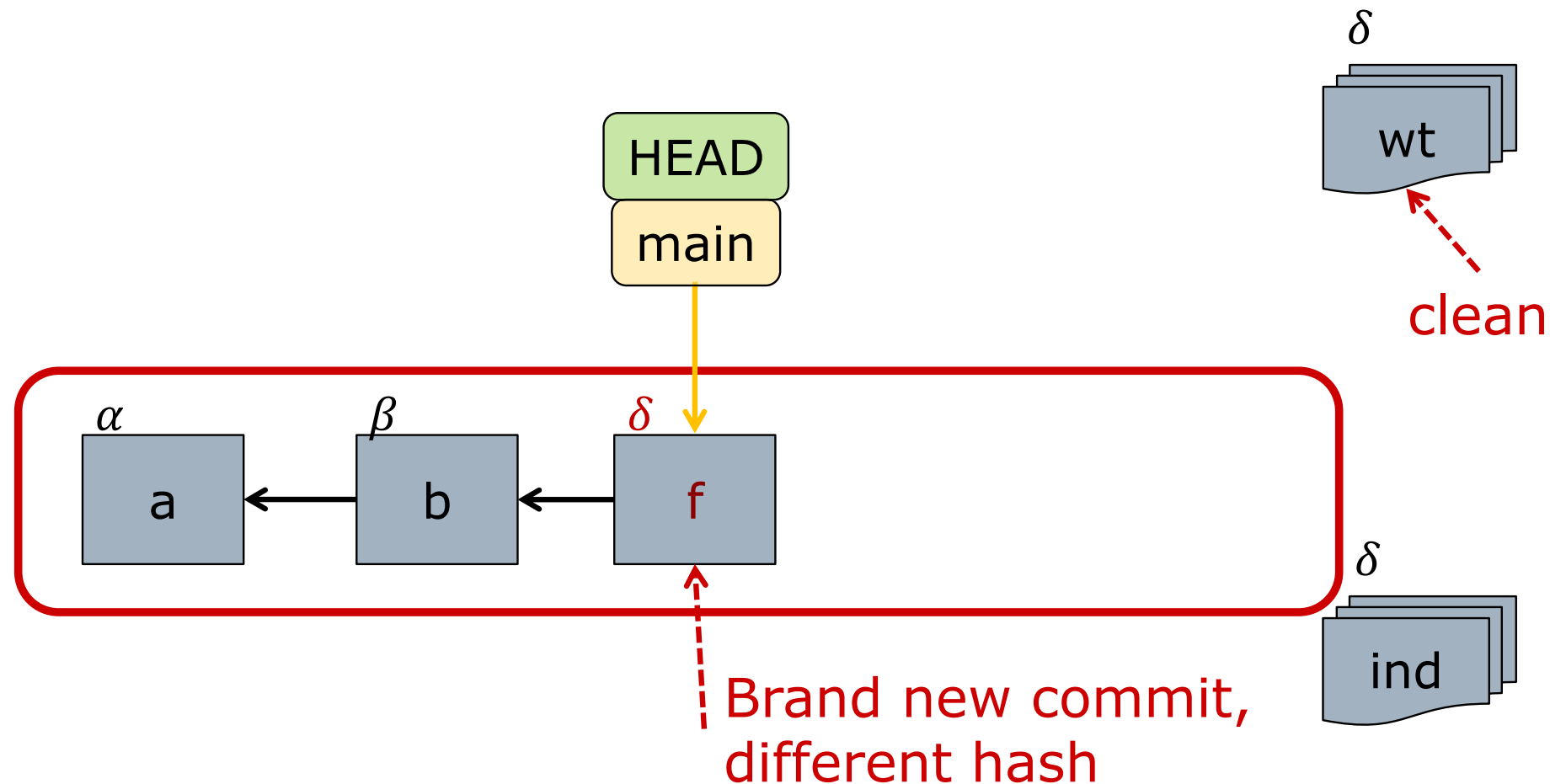
Soln: Repair Tip with Commit --amend

- Alternative: Change most recent commit(s)



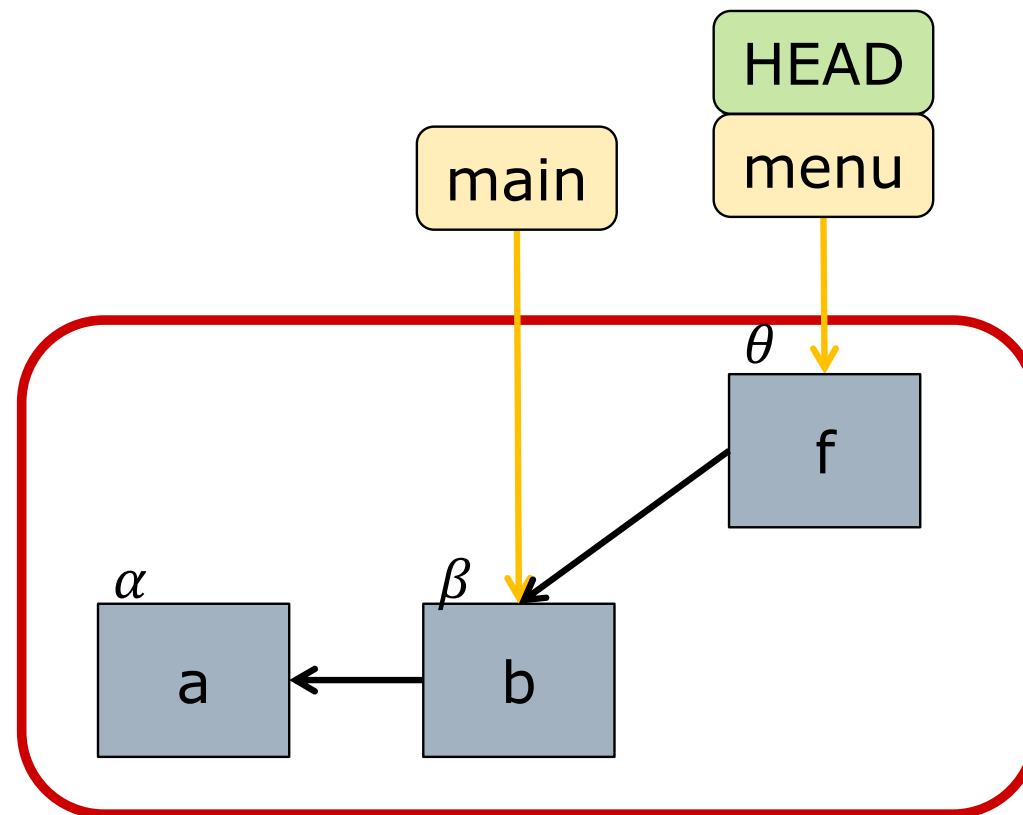
Repair Tip with Commit -amend (After)

```
$ git add .  
$ git commit --amend --no-edit  
# no-edit means keep the same commit message
```



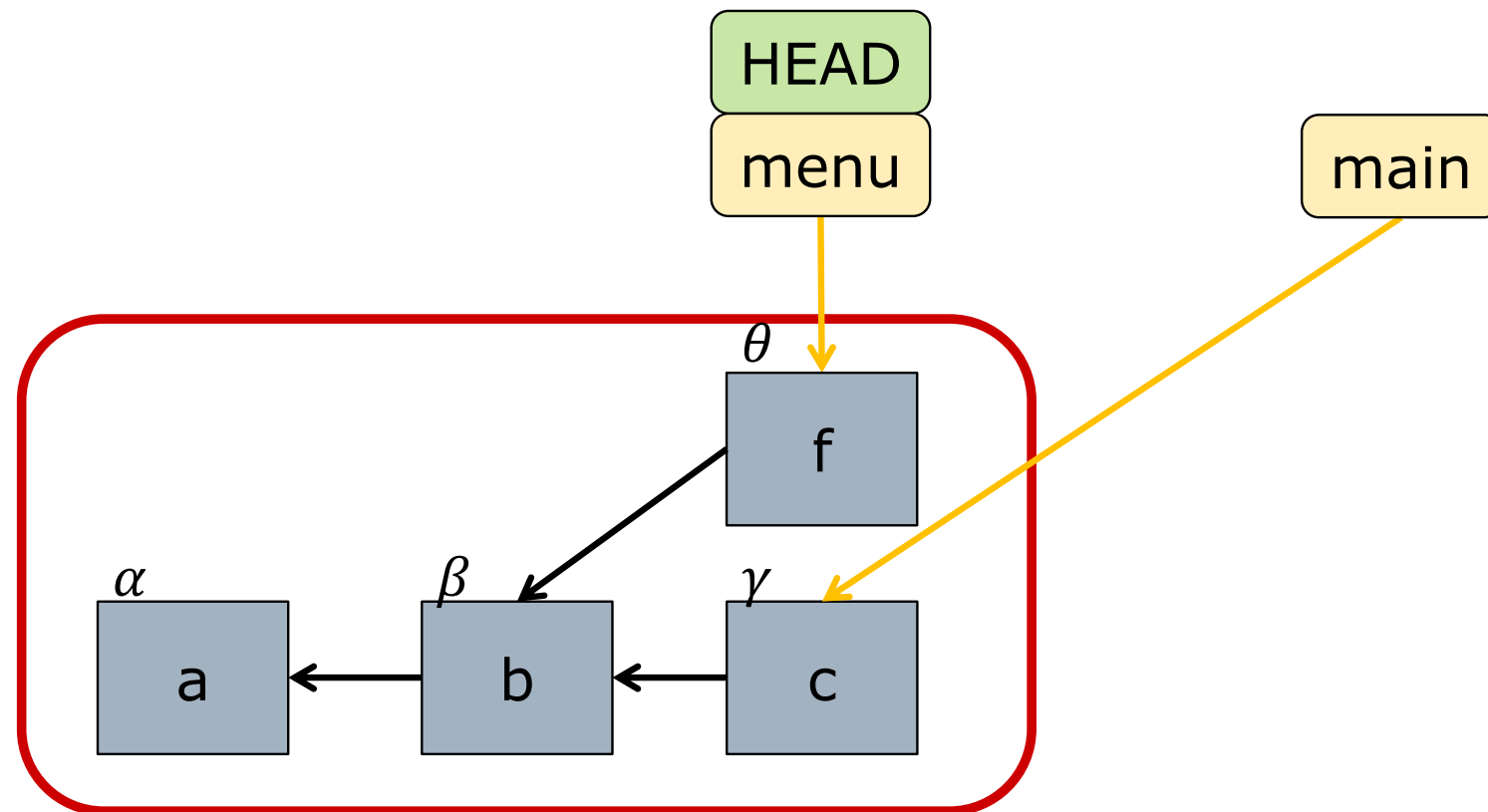
Problem 2: Periodic Merging (1)

- Problem 2: As an independent branch is being developed, main also evolves



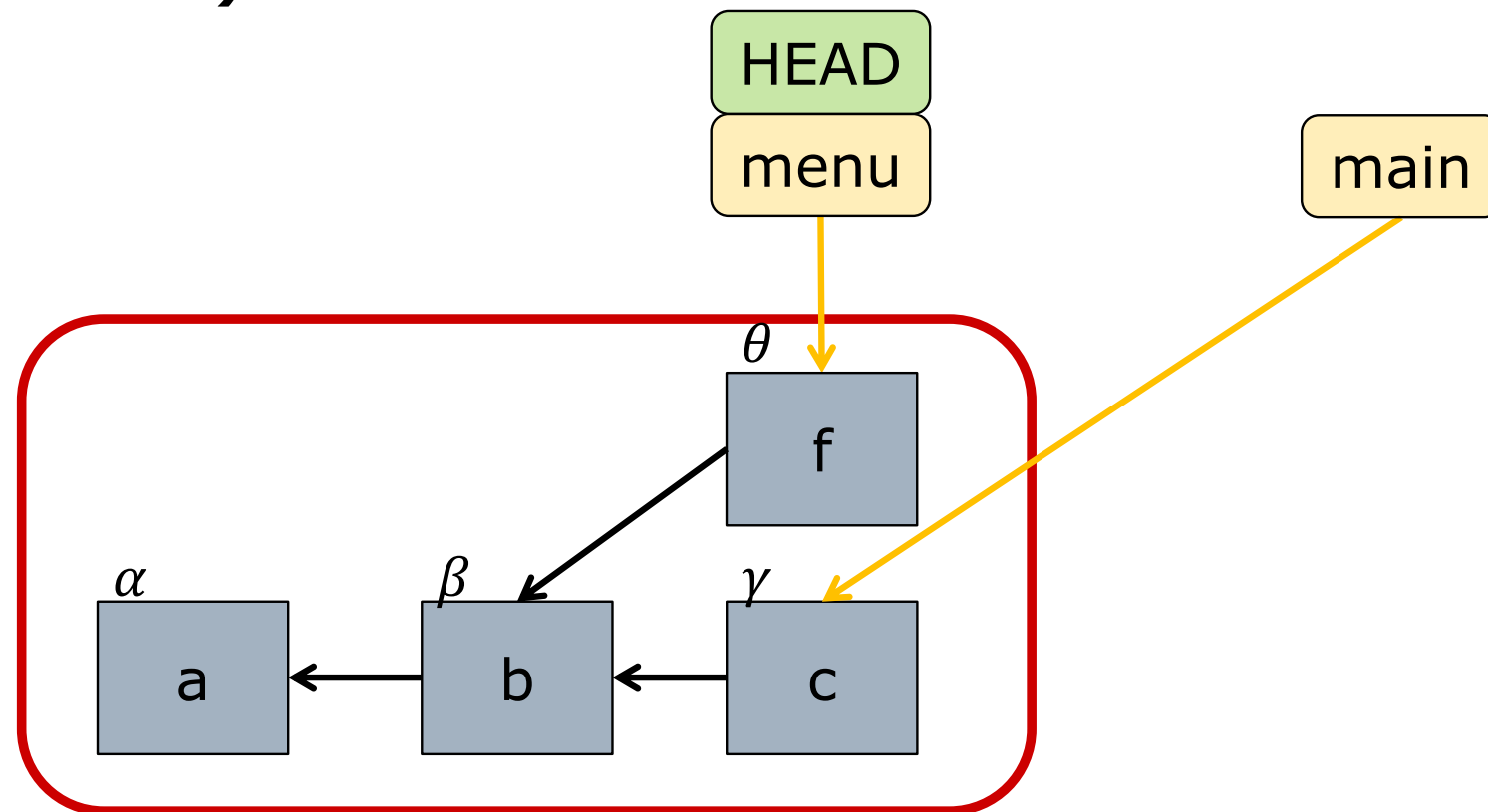
Problem 2: Periodic Merging (2)

- Problem 2: As an independent branch is being developed, main also evolves



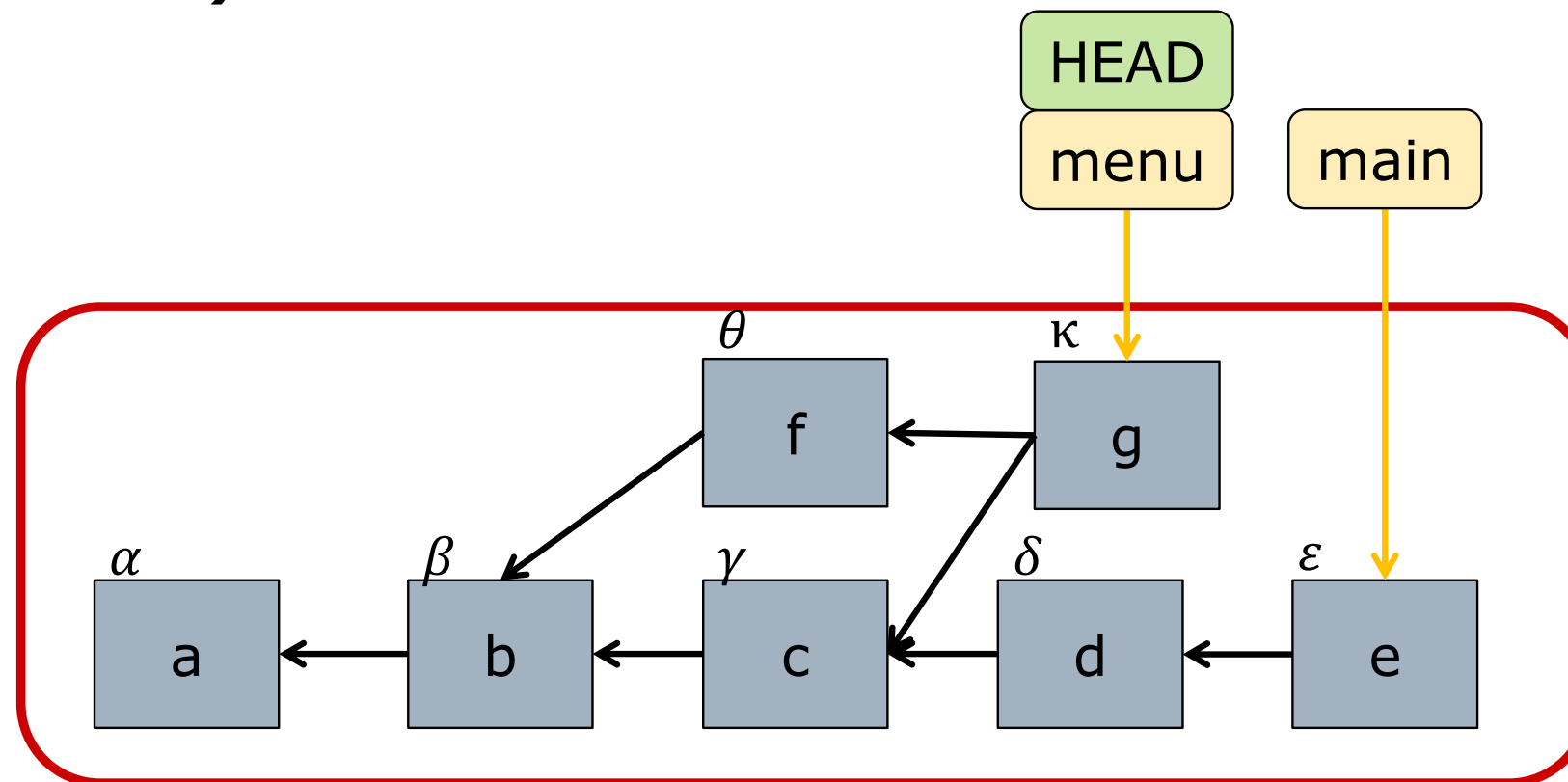
Problem 2: Periodic Merging (3)

- ❑ Problem 2: As an independent branch is being developed, main also evolves
- ❑ Result: Need periodic merges of main with (incomplete) branch



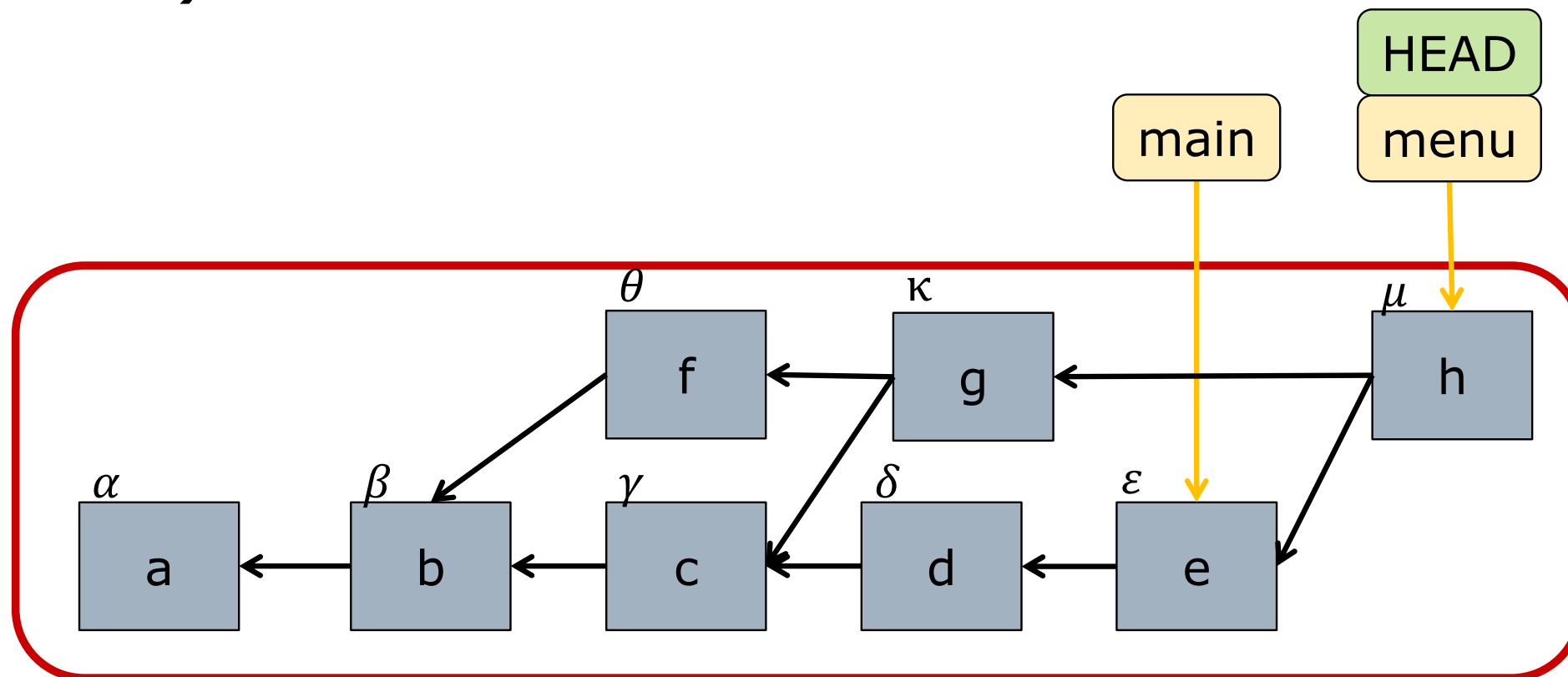
Problem 2: Periodic Merging (4)

- ❑ Problem 2: As an independent branch is being developed, main also evolves
- ❑ Result: Need periodic merges of main with (incomplete) branch



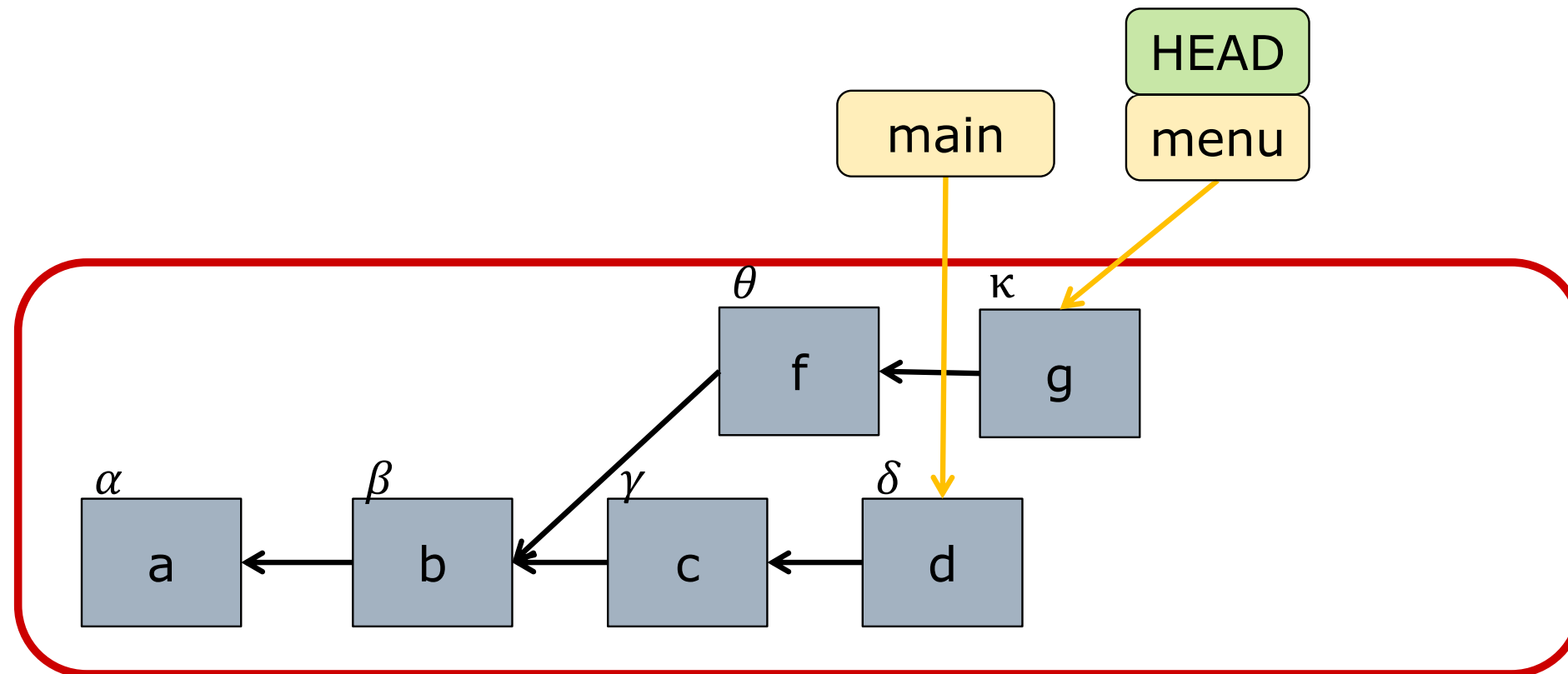
Problem 2: Periodic Merging (5)

- ❑ Problem 2: As an independent branch is being developed, main also evolves
- ❑ Result: Need periodic merges of main with (incomplete) branch



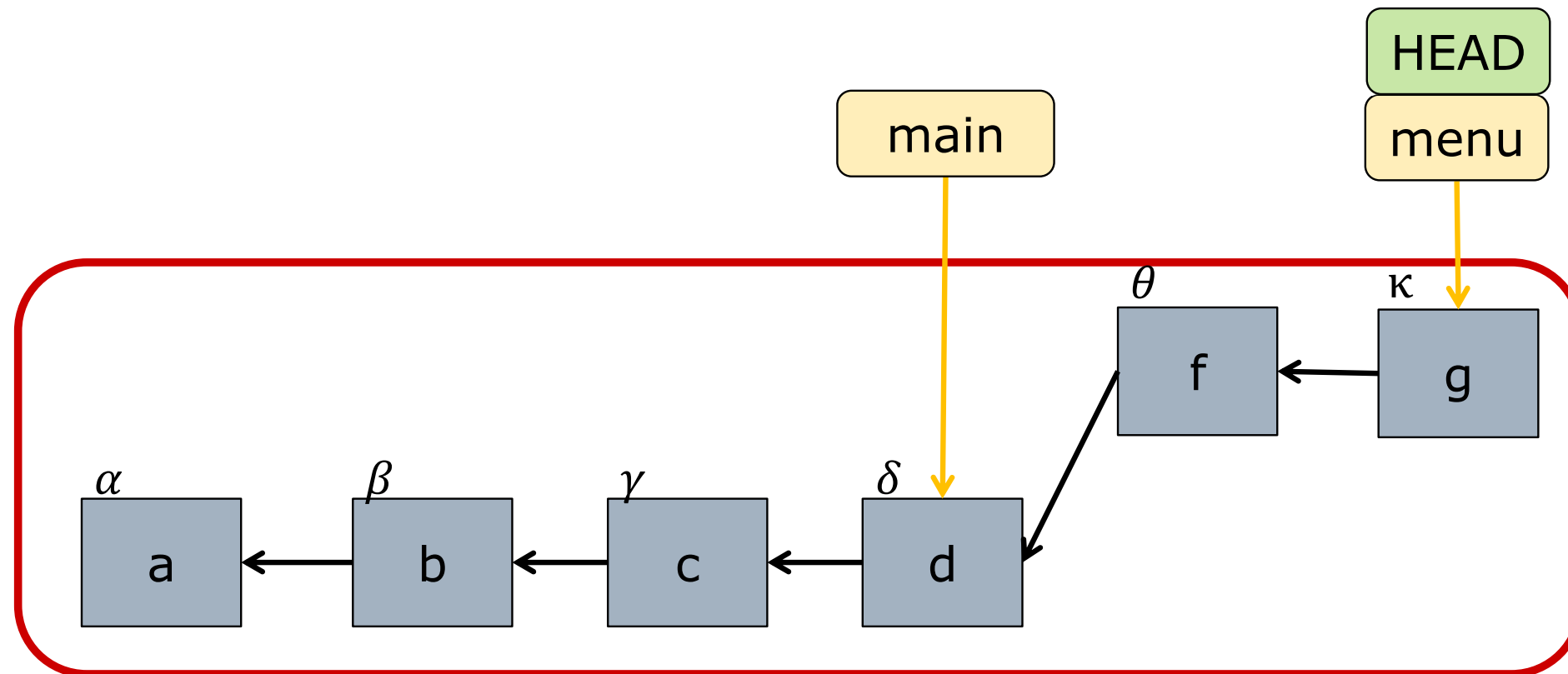
Soln: DAG Surgery with Rebase

- Alternative: Move commits to a different part of the DAG



Soln: DAG Surgery with Rebase (After)

```
$ git rebase main  
# merging menu into main is now a fast-forward
```



Git Clients and Hosting Services

- ❑ Recommend'n: Know the command line!
- ❑ IDEs are helpful too
 - VSCode, plus Git Graph extension
- ❑ Lots of sites for hosting your repos:
 - GitHub, GitLab, Bitbucket, SourceForge...
 - See: git.wiki.kernel.org/index.php/GitHosting
- ❑ These cloud services provide
 - Storage space, account/access management
 - Pretty web interface
 - Issues, bug tracking
 - Workflow (eg forks) to promote contributions from others

Clarity

git != GitHub



Warning: Academic Misconduct

- GitHub is a very popular service
 - New repos are *public* by default
 - Even free plan allows unlimited *private* repo's (and collaborators)
 - 3901 has an organization for your private repo's and team access
- Other services (*e.g.* GitLab, Bitbucket) have similar issues
- Public repo's containing coursework can create academic misconduct issues
 - Problems for poster
 - Problems for plagiarist

Summary: Advanced Git

- Workflow
 - Fetch/push frequency
 - Respect team conventions for how/when to use different branches
- Central repo is a shared resource
 - Contains common (source) code
 - Normalize line endings and formats
- Advanced techniques
 - Stash, reset, rebase
- Advice
 - Learn by using the command line
 - Beware academic misconduct