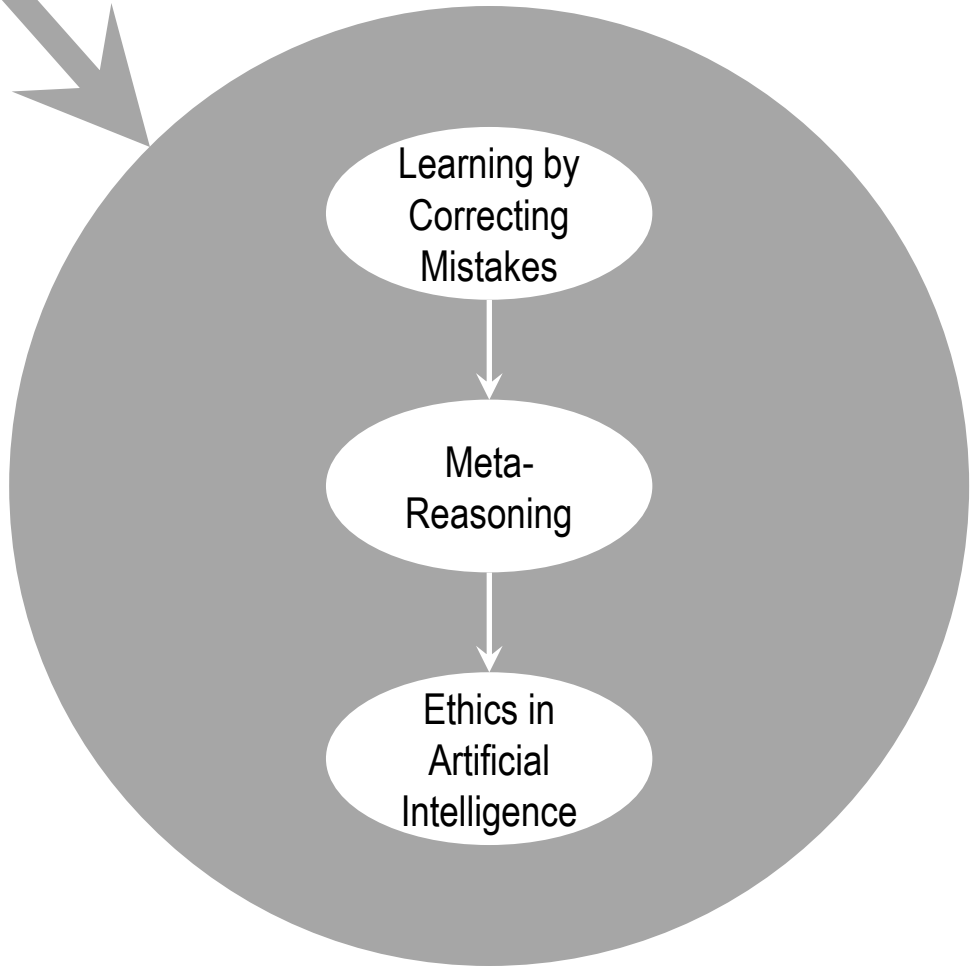
Learning by Correcting Mistakes

Metacognition



Lesson Preview

- Explanation-based learning revisited
- Isolating mistakes
- Explaining mistakes
- Correcting mistakes

A Cup

A cup is an object that is stable and enables drinking.

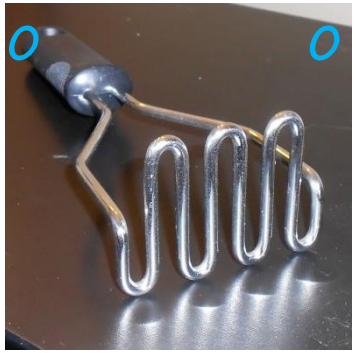
An Object

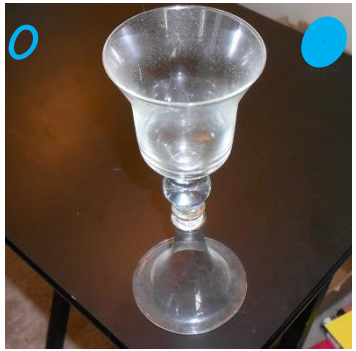
This object is light and made of porcelain. It has a decoration, a concavity, and a handle. The bottom is flat.

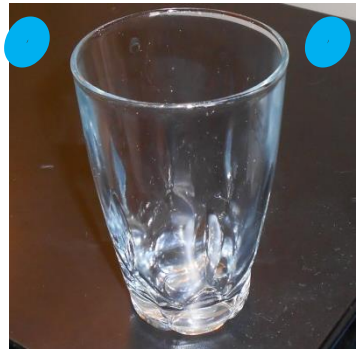
Can we prove this object is a cup?

Which of these objects are cups? Check the box in the top left for each cup.

Which of these objects meet our definition of a cup? Check the box in the top right for each object that meets our definition.







A Cup

A cup is an object that is stable and enables drinking.

An Object

This object is light and made of porcelain. It has a decoration, a concavity, and a handle. The bottom is flat.

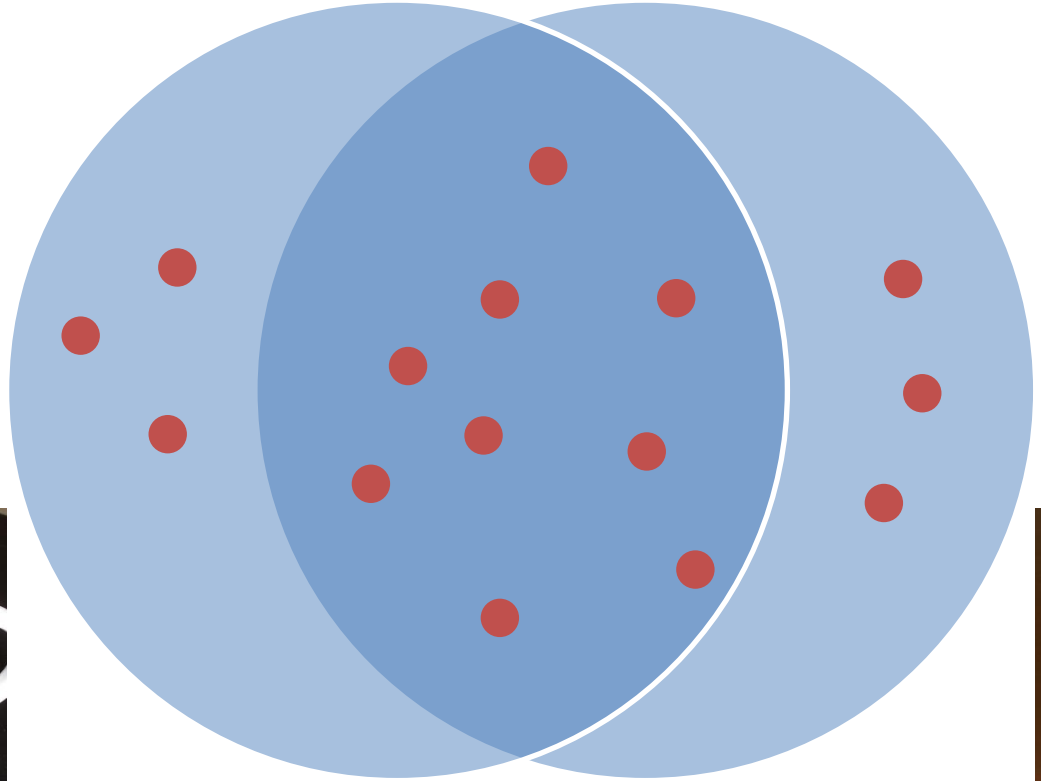
Can we prove this object is a cup?

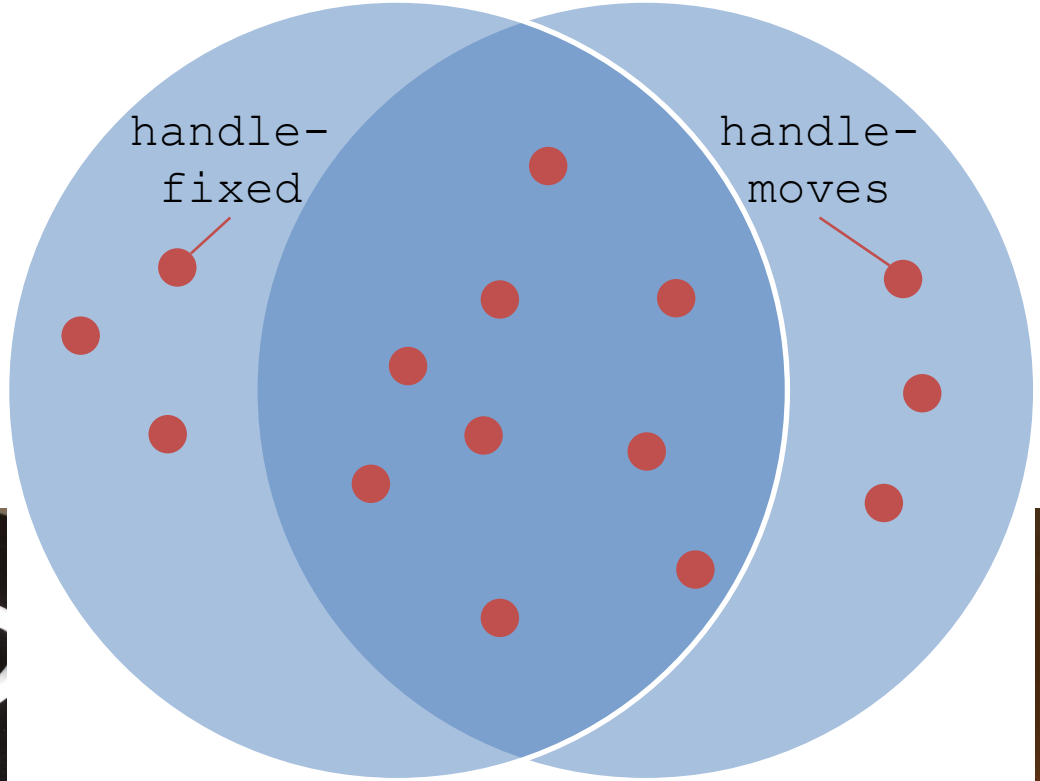
Questions for Learning from Mistakes

1. How can the agent isolate the error in its former model?

2. How can the agent explain the problem that led to the error?

3. How can the agent repair the model to prevent the error from recurring?




handle-
fixed
handle-
moves

Algorithm for Isolating Mistakes

To find suspicious true-success relations:
- Intersect all true successes (∩T)
- Union all false successes (∪F)
- Remove assertions in union from intersection (∩T - ∪F)

To find suspicious false-success relations:
- Intersect all false successes (∩F)
- Union all true successes (∪T)
- Remove all assertions in union from intersection (∩F - ∪T)

Old Rule

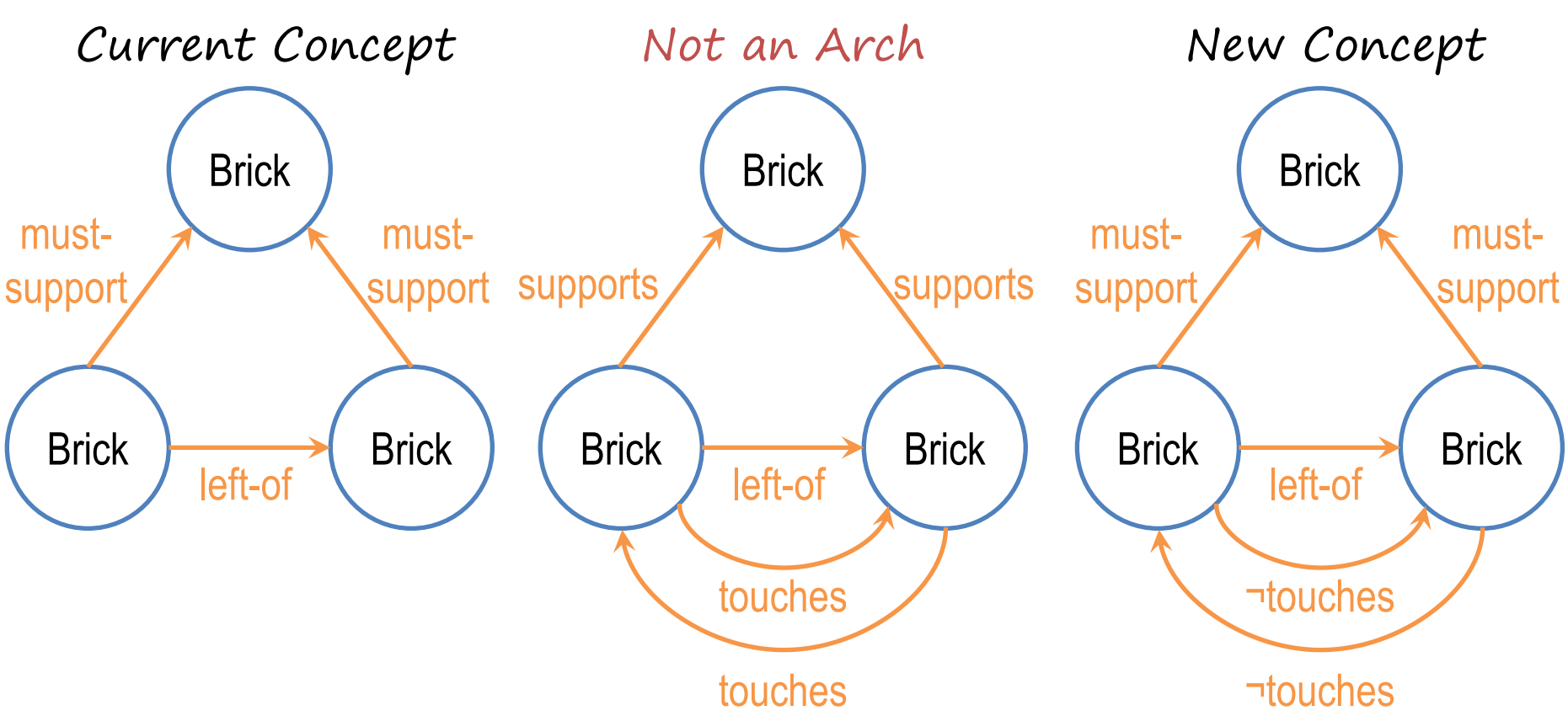
If:
- Object has bottom
- Bottom is flat
- Object has concavity
- Object is lightweight
- Object has a handle

Then:
- Object is a cup

New Rule

If:
- Object has bottom
- Bottom is flat
- Object has concavity
- Object is lightweight
- Object has a handle
- Handle is fixed

Then:
- Object is a cup

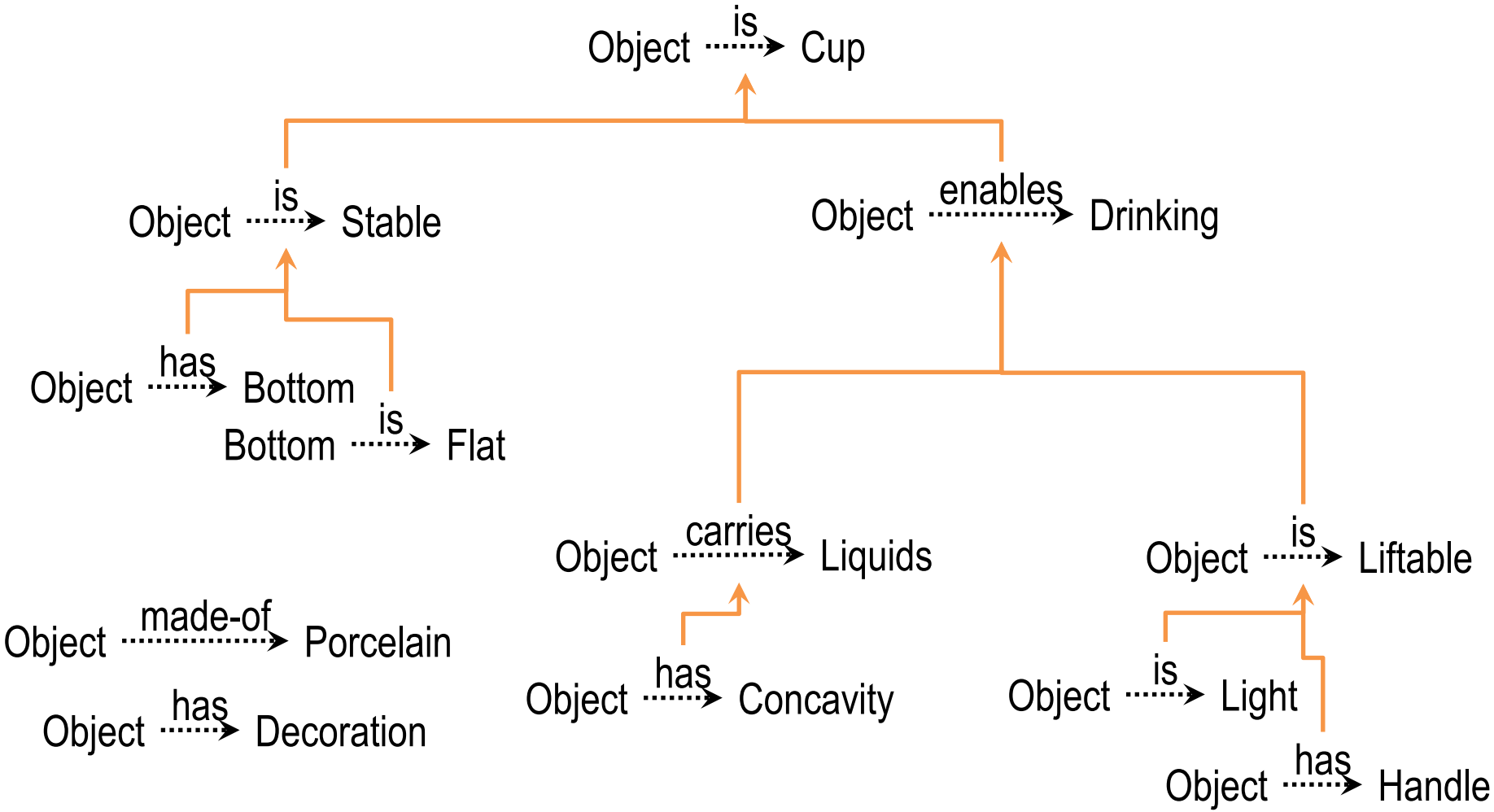

Current Concept
Brick
must-support
must-support
Brick
left-of
Brick
Not an Arch
Brick
supports
supports
Brick
left-of
Brick
touches
touches
New Concept
Brick
must-support
must-support
Brick
left-of
Brick
¬touches
¬touches

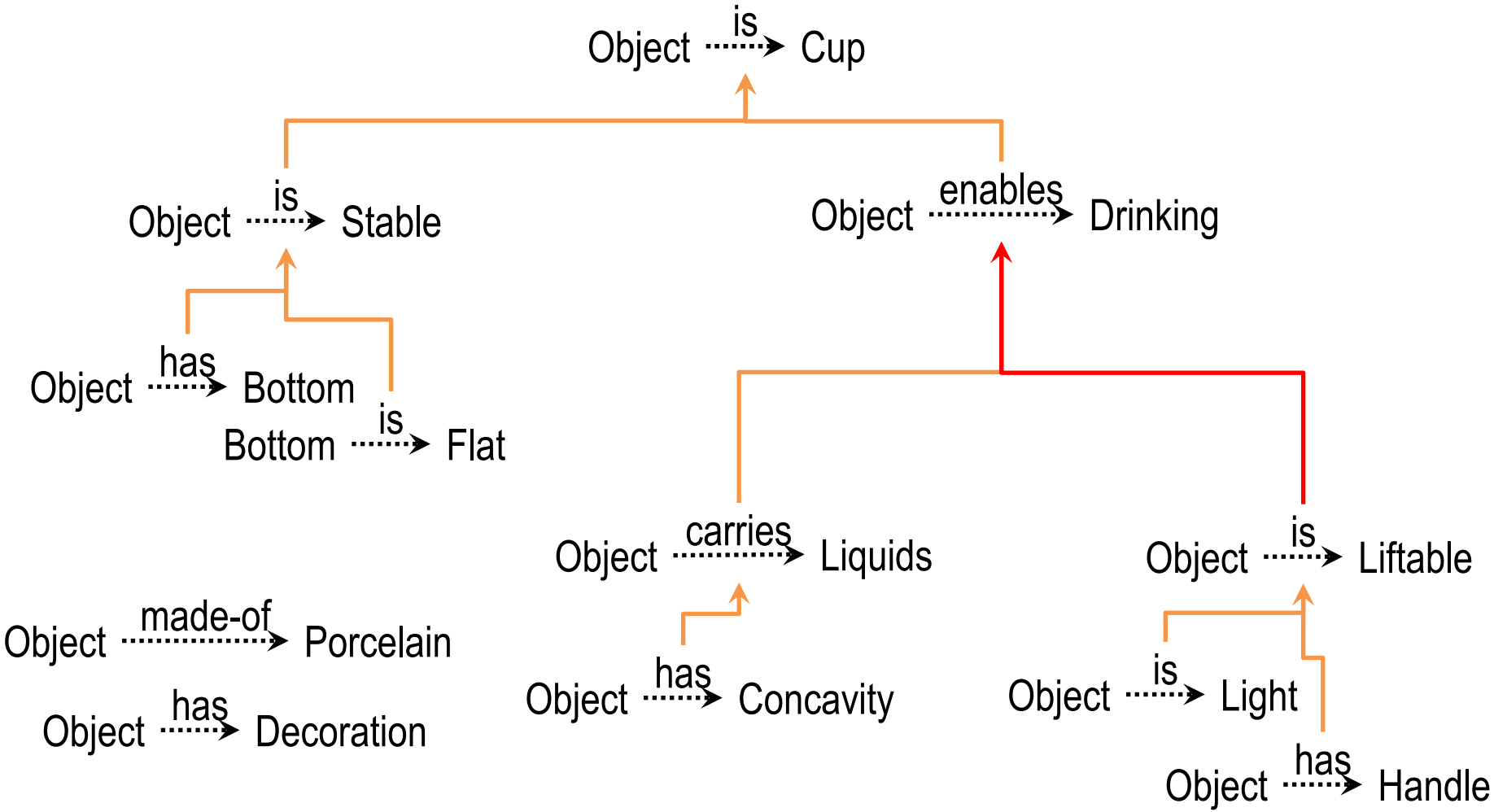
Questions for Learning from Mistakes

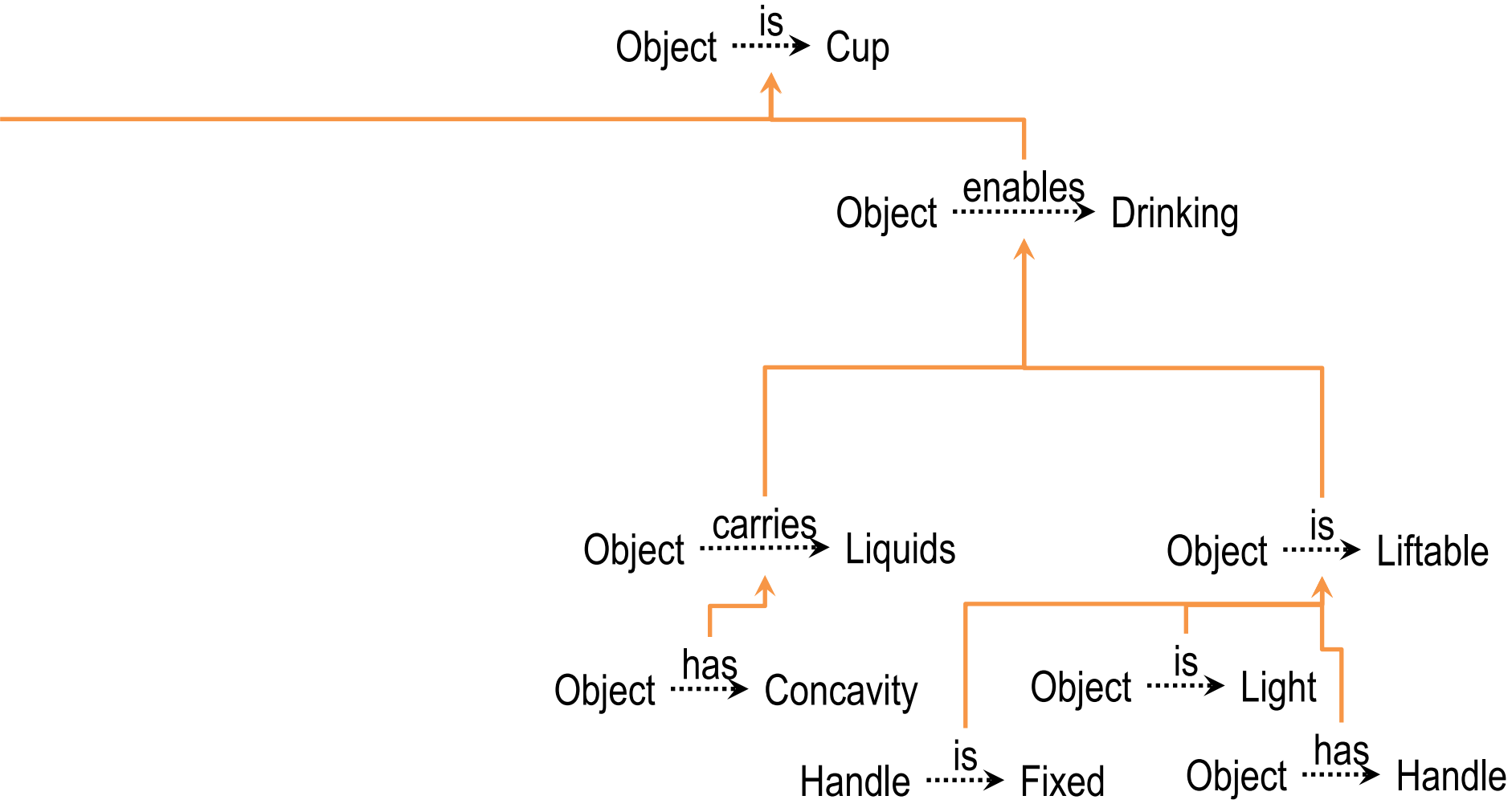
1. How can the agent isolate the error in its former model?

2. How can the agent explain the problem that led to the error?

3. How can the agent repair the model to prevent the error from recurring?

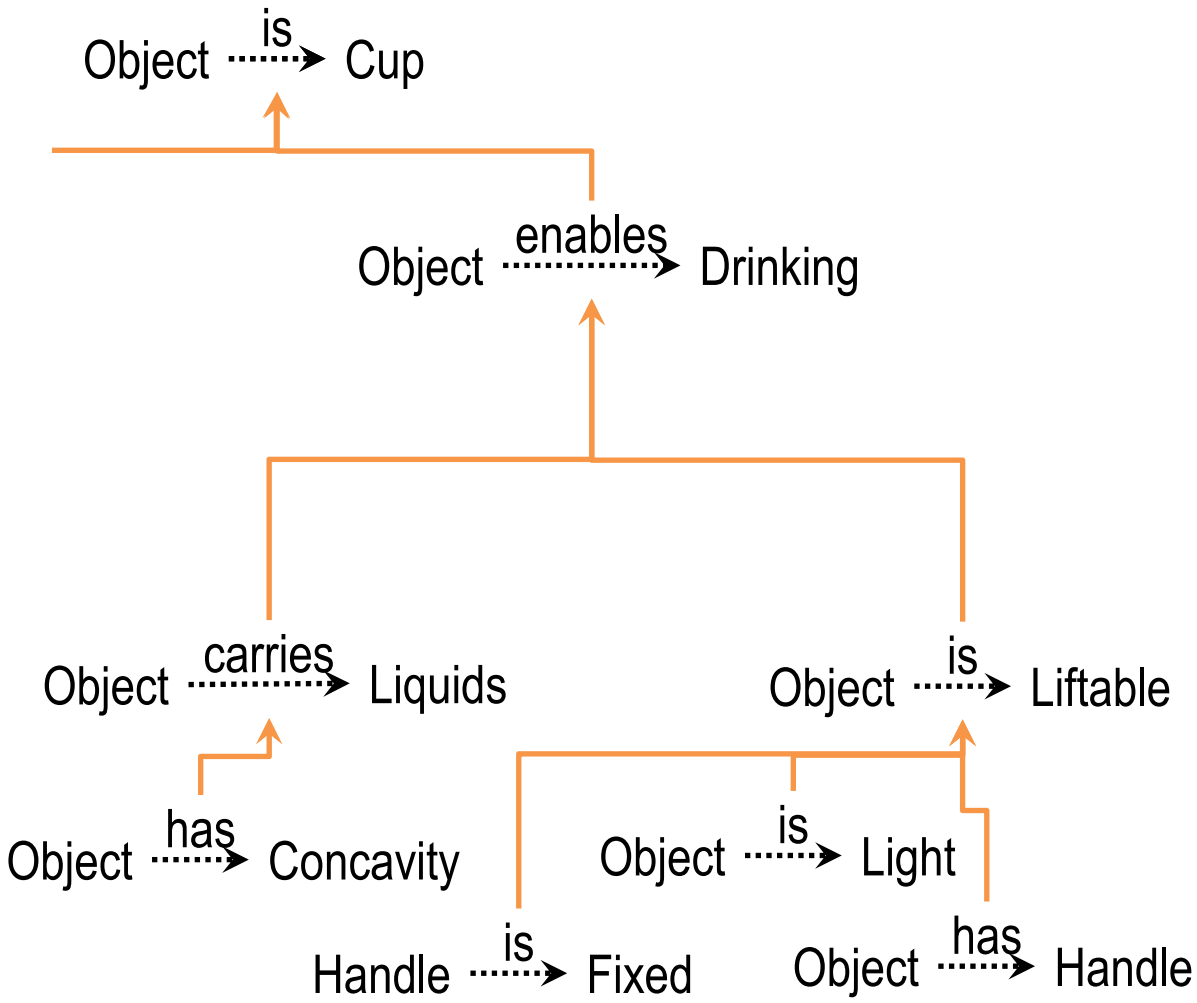

Object is Cup
Object is Stable
Object enables Drinking
Object has Bottom
Bottom is Flat
Object made-of Porcelain
Object has Decoration
Object carries Liquids
Object has Concavity
Object is Liftable
Object is Light
Object has Handle

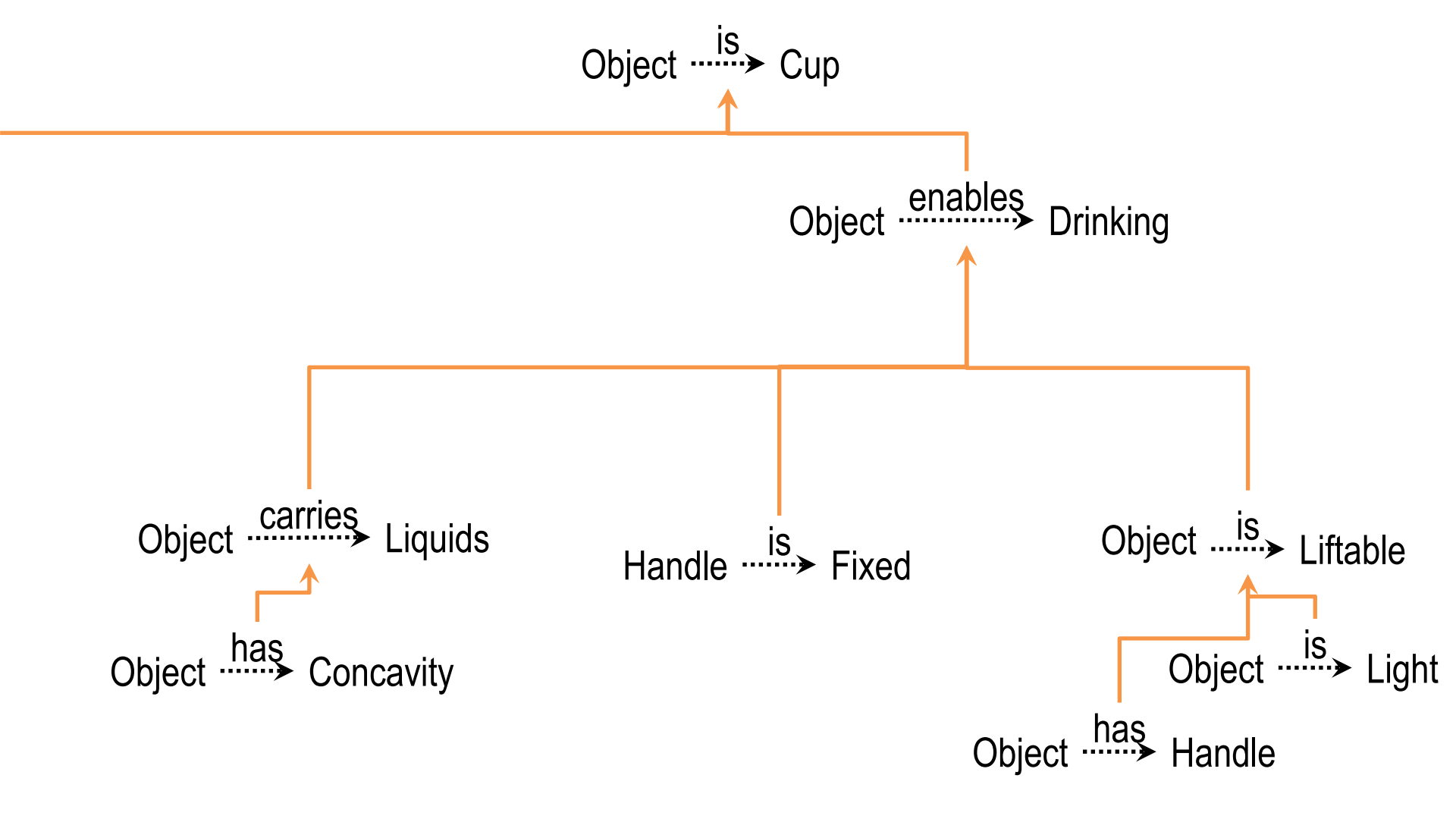

Object is Cup
Object is Stable
Object enables Drinking
Object has Bottom
Bottom is Flat
Object made-of Porcelain
Object has Decoration
Object carries Liquids
Object has Concavity
Object is Liftable
Object is Light
Object has Handle

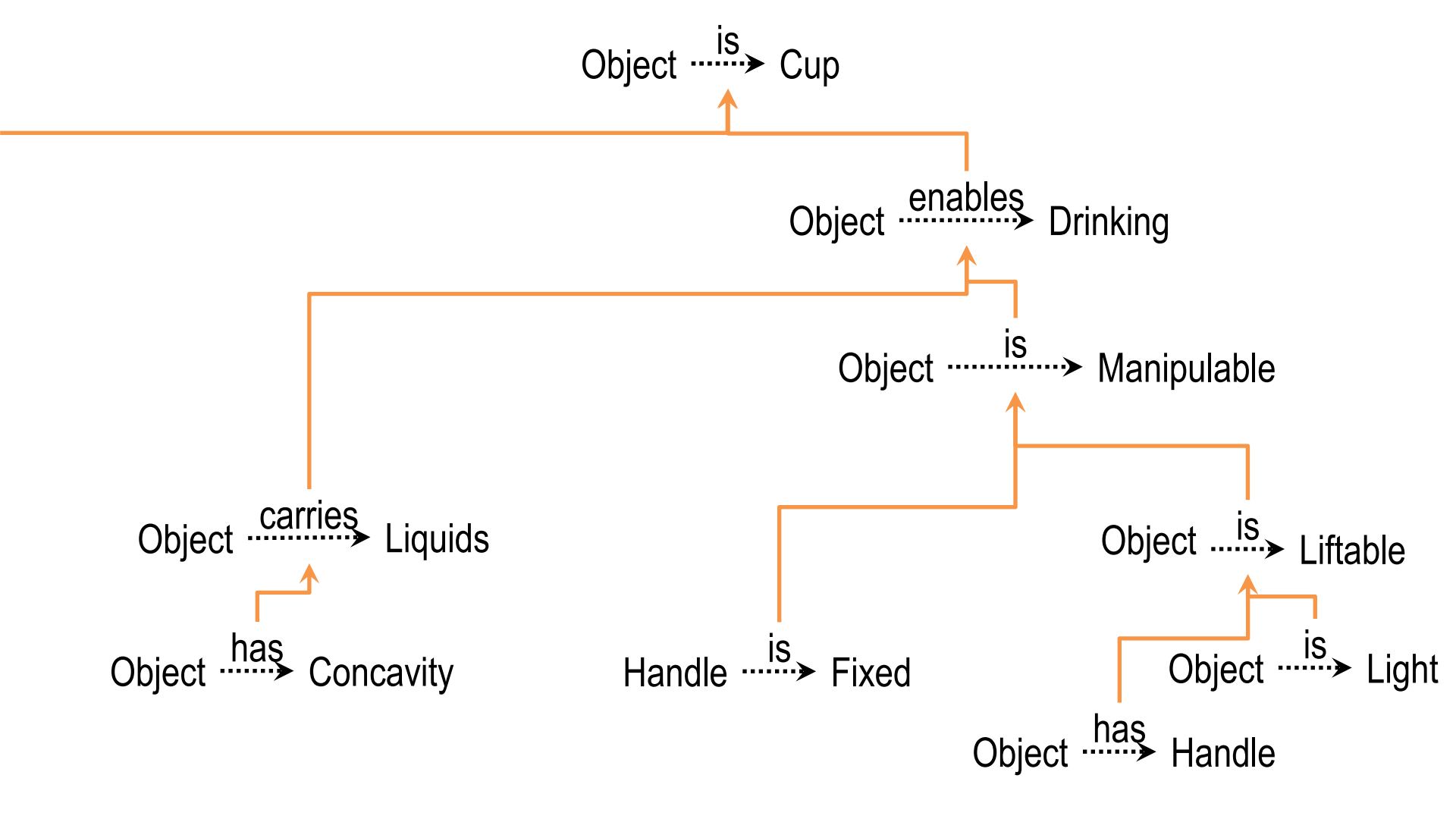

Object is Cup
Object enables Drinking
Object carries Liquids
Object is Liftable
Object has Concavity
Object is Light
Handle is Fixed
Object has Handle

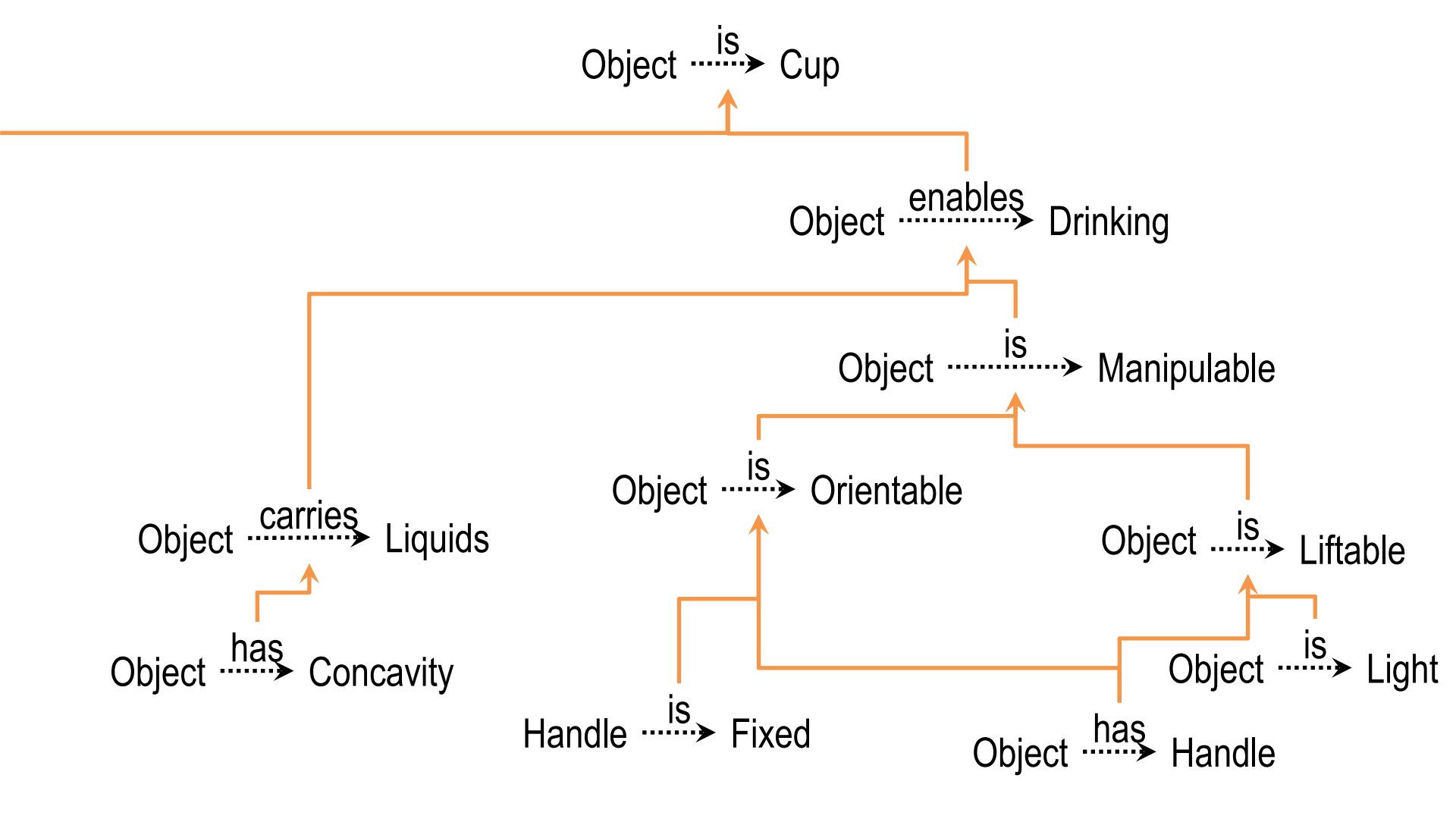
Is this a good way to fix this error?

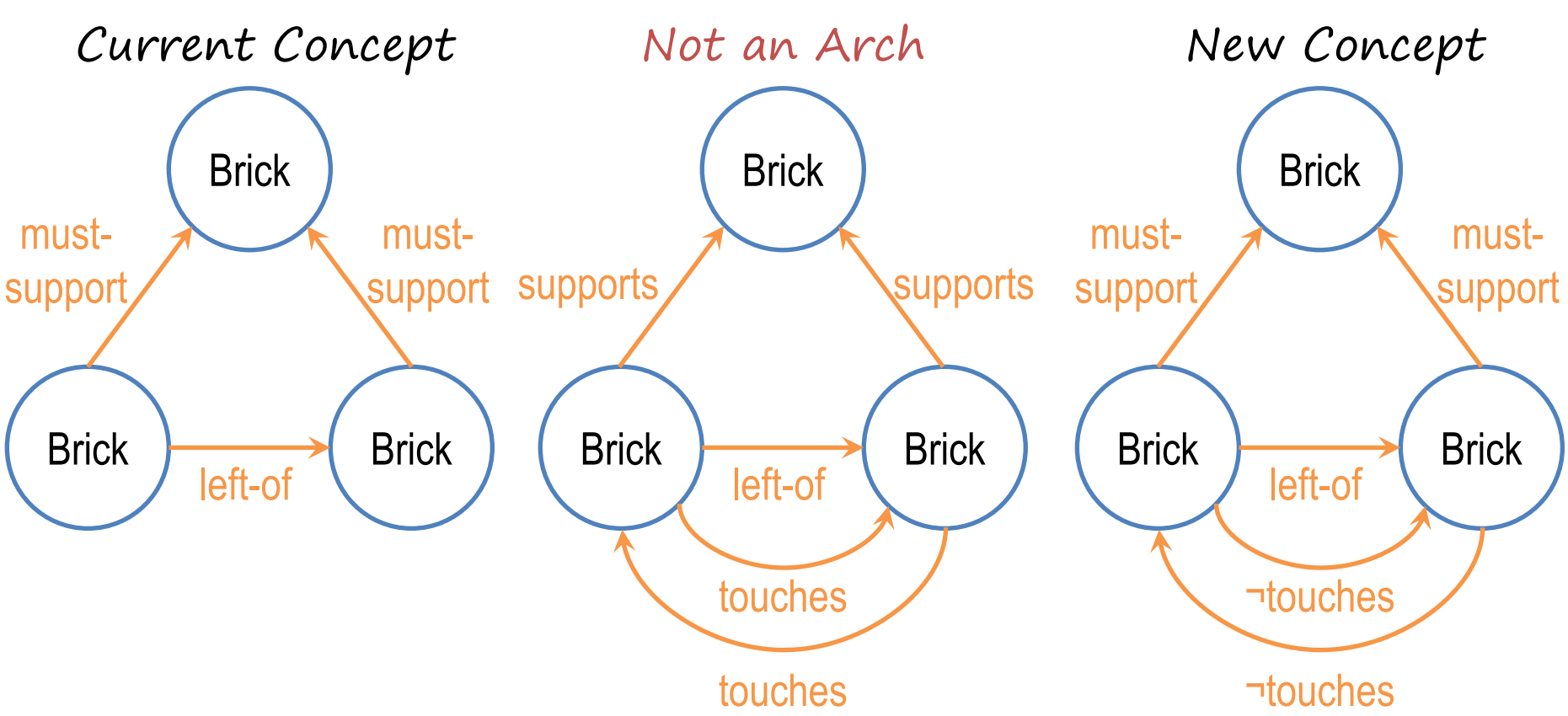
- Yes, because it shows only fixed-handle cups enable drinking.
- No, because it will exclude some actual cups.
- No, because some non-cups will still be included.
- No, because it will cause incorrect decisions about other objects.




Object is Cup
Object enables Drinking
Object carries Liquids
Handle is Fixed
Object is Liftable
Object has Concavity
Object is Light
Object has Handle


Object is Cup
Object enables Drinking
Object is Manipulable
Object carries Liquids
Object has Concavity
Handle is Fixed
Object is Liftable
Object is Light
Object has Handle


Object is Cup
Object enables Drinking
Object is Manipulable
Object is Orientable
Object carries Liquids
Object is Liftable
Object has Concavity
Object is Light
Handle is Fixed
Object has Handle


Current Concept
Brick
must-support
must-support
Brick
left-of
Brick
Not an Arch
Brick
supports
supports
Brick
left-of
Brick
touches
touches
New Concept
Brick
must-support
must-support
Brick
left-of
Brick
¬touches
¬touches

Assignment

How would you use diagnosis to design an agent that could answer Raven's progressive matrices?

To recap…

- Explanation-based learning and incremental concept learning revisited
- Isolating mistakes
- Explaining mistakes
- Correcting mistakes