



# Eliminative Argumentation: A Method for Assessing Confidence in a Claim

John B. Goodenough

Charles B. Weinstock

Software Engineering Institute  
Carnegie Mellon University  
Pittsburgh, PA 15213

# Document Markings

Copyright 2018 Carnegie Mellon University. All Rights Reserved.

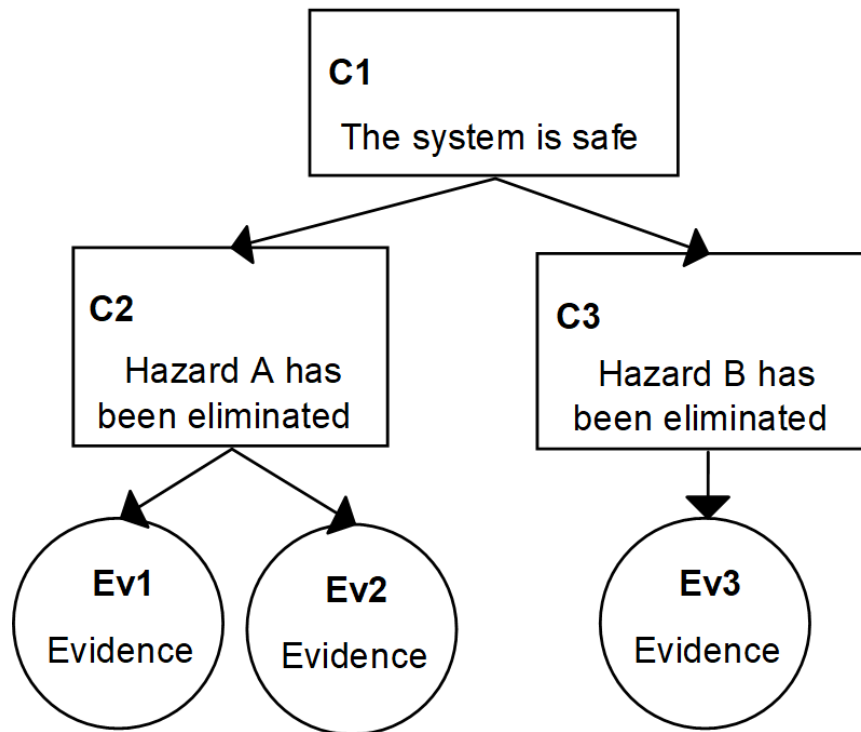
NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other use. Requests for permission should be directed to the Software Engineering Institute at [permission@sei.cmu.edu](mailto:permission@sei.cmu.edu).

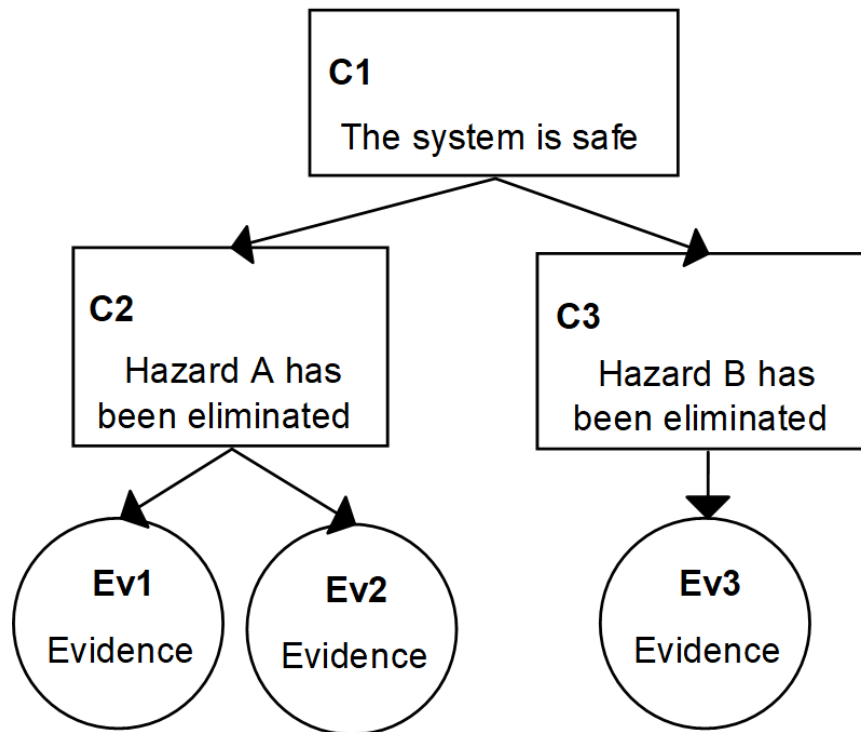
DM18-0181

# A Notional Assurance Case



# The Problem

Given the evidence, how confident should we be in the claim C1? Why?  
What does it mean to have confidence in the claim?  
What could be done to improve confidence? Why?



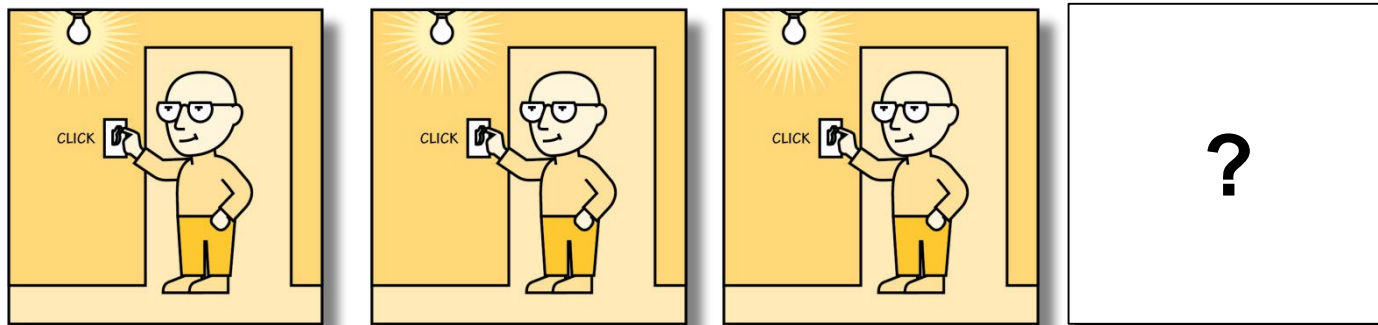
# The Basis for Confidence in a Claim

A classic philosophical problem:

- Justify belief in a hypothesis

Use Induction

- Enumerative: Support increases as **confirming instances** are found



**Using past experience as the basis for predicting future behavior**

# The Basis for Confidence in a Claim

A classic philosophical problem:

- Justify belief in a hypothesis

Use Induction

- **Eliminative**: Support increases as **reasons for doubt** are eliminated
  - Switch not connected to light
  - No power
  - Dead light bulb

# The Basis for Confidence in a Claim

A classic philosophical problem:

- Justify belief in a hypothesis

Use Induction

- **Eliminative**: Support increases as **reasons for doubt** are eliminated
  - Switch not connected to light
  - No power
  - Dead light bulb



# The Basis for Confidence in a Claim

A classic philosophical problem:

- Justify belief in a hypothesis

Use Induction

- **Eliminative**: Support increases as **reasons for doubt** are eliminated
  - ~~Switch not connected to light~~
  - No power
  - Dead light bulb



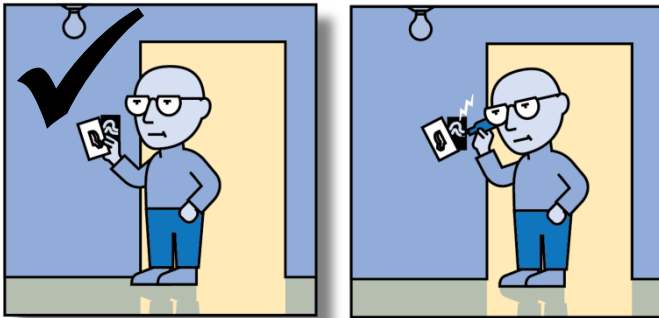
# The Basis for Confidence in a Claim

A classic philosophical problem:

- Justify belief in a hypothesis

Use Induction

- **Eliminative**: Support increases as **reasons for doubt** are eliminated
  - ~~Switch not connected to light~~
  - **No power**
  - Dead light bulb



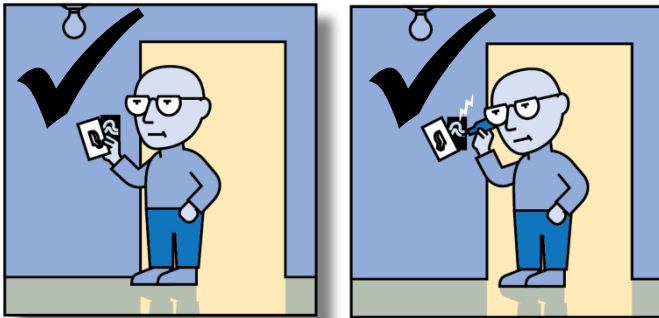
# The Basis for Confidence in a Claim

A classic philosophical problem:

- Justify belief in a hypothesis

Use Induction

- **Eliminative**: Support increases as **reasons for doubt** are eliminated
  - ~~Switch not connected to light~~
  - ~~No power~~
  - Dead light bulb



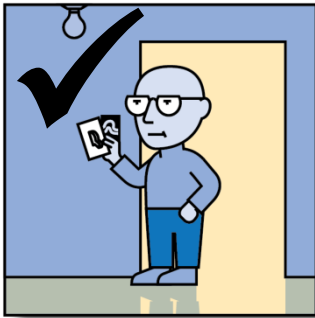
# The Basis for Confidence in a Claim

A classic philosophical problem:

- Justify belief in a hypothesis

Use Induction

- **Eliminative**: Support increases as **reasons for doubt** are eliminated
  - ~~Switch not connected to light~~
  - ~~No power~~
  - **Dead light bulb**



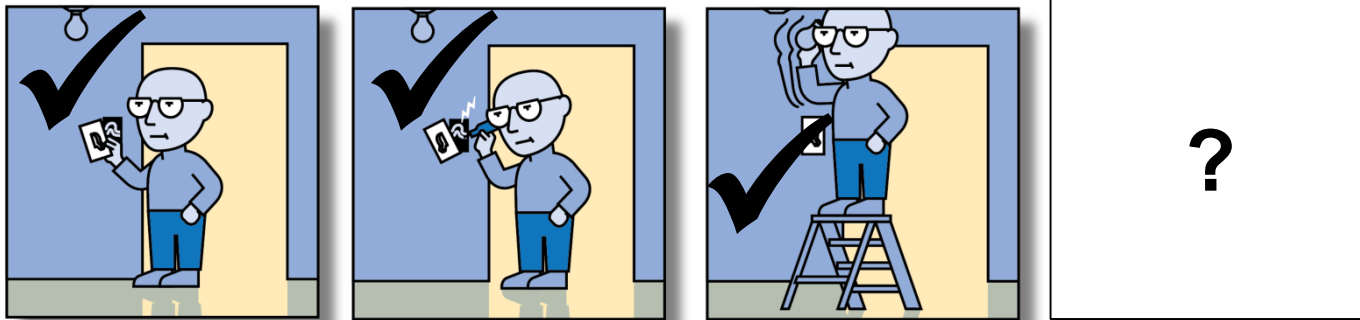
# The Basis for Confidence in a Claim

A classic philosophical problem:

- Justify belief in a hypothesis

Use Induction

- **Eliminative**: Support increases as **reasons for doubt** are eliminated
  - ~~Switch not connected to light~~
  - ~~No power~~
  - ~~Dead light bulb~~



# The Basis for Confidence in a Claim

A classic philosophical problem:

- Justify belief in a hypothesis

Use Induction

- **Eliminative**: Support increases as **reasons for doubt** are eliminated
  - ~~Switch not connected to light~~ – **3/3 (complete confidence)**
  - ~~No power~~
  - ~~Dead light bulb~~



# The Basis for Confidence in a Claim

A classic philosophical problem:

- Justify belief in a hypothesis

Use Induction

- **Eliminative**: Support increases as **reasons for doubt** are eliminated
  - Switch not connected to light
  - No power
  - Dead light bulb
  - 3/3 (complete confidence)
  - 0/3 (no confidence)



# The Basis for Confidence in a Claim

A classic philosophical problem:

- Justify belief in a hypothesis

Use Induction

- **Eliminative**: Support increases as **reasons for doubt** are eliminated
    - Switch not connected to light
    - ~~No power~~
    - ~~Dead light bulb~~
- 3/3 (complete confidence)  
– 0/3 (no confidence)  
– **2/3 (partial confidence)**

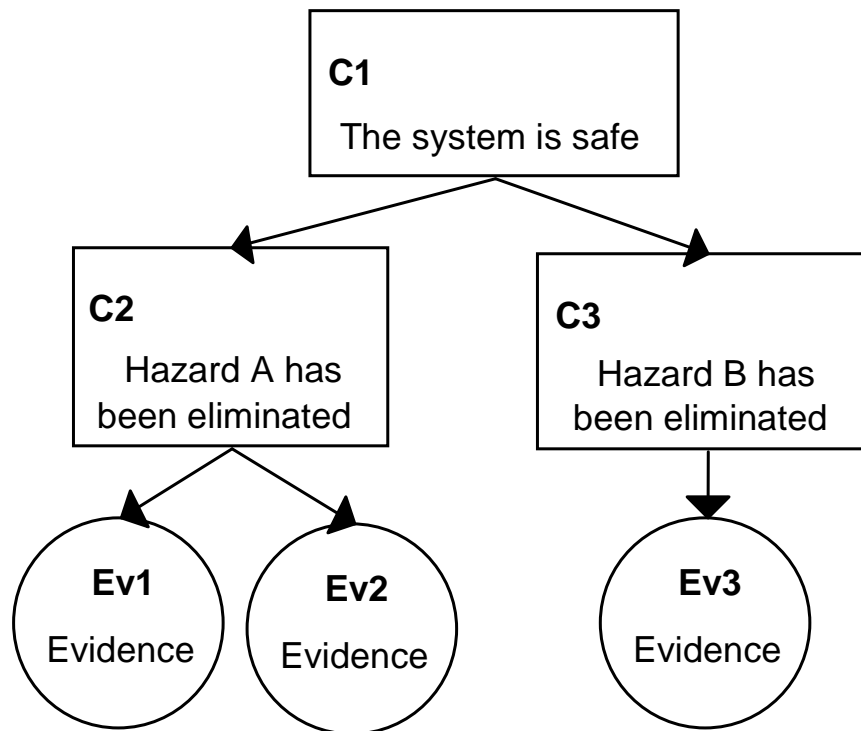


# The Problem

How confident in C1? Why? (Number of uneliminated doubts)

What does it mean to have confidence? (To have no doubt)

What could be done to improve confidence? Why? (Elim. more doubts)



# Eliminative Argumentation

A form of structured argumentation incorporating concepts from

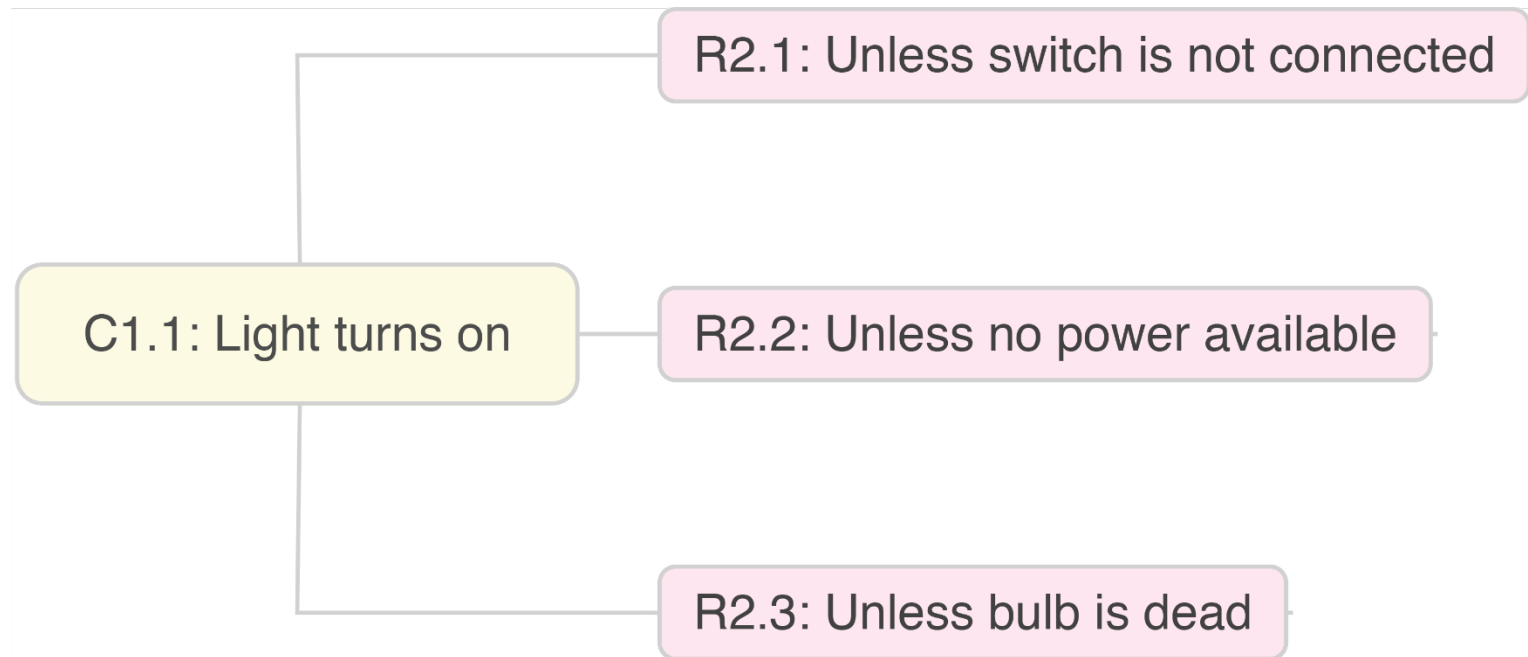
- Assurance cases
- Eliminative induction
- Defeasible reasoning

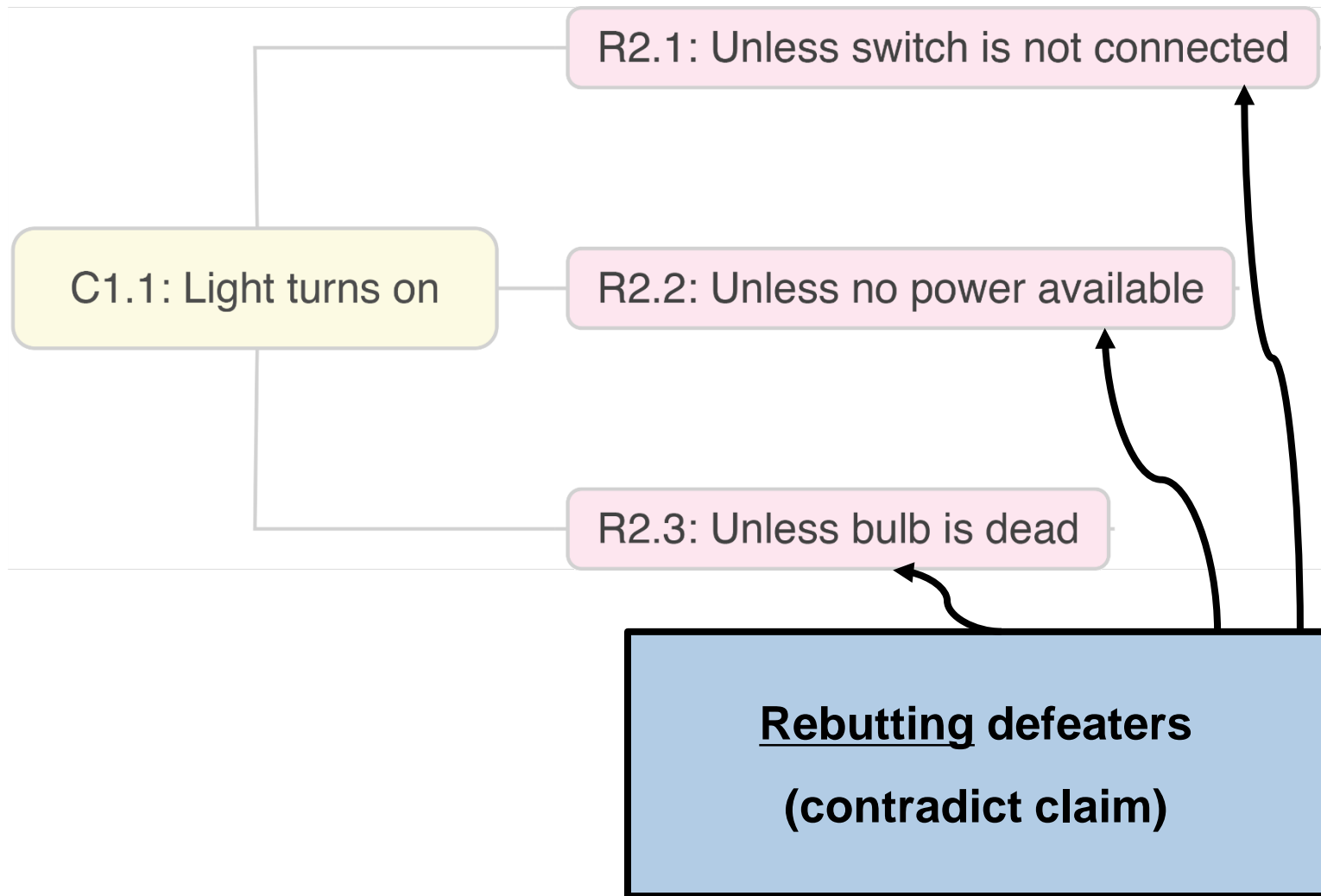
**An eliminative argument shows the basis for confidence in an argument's conclusion**

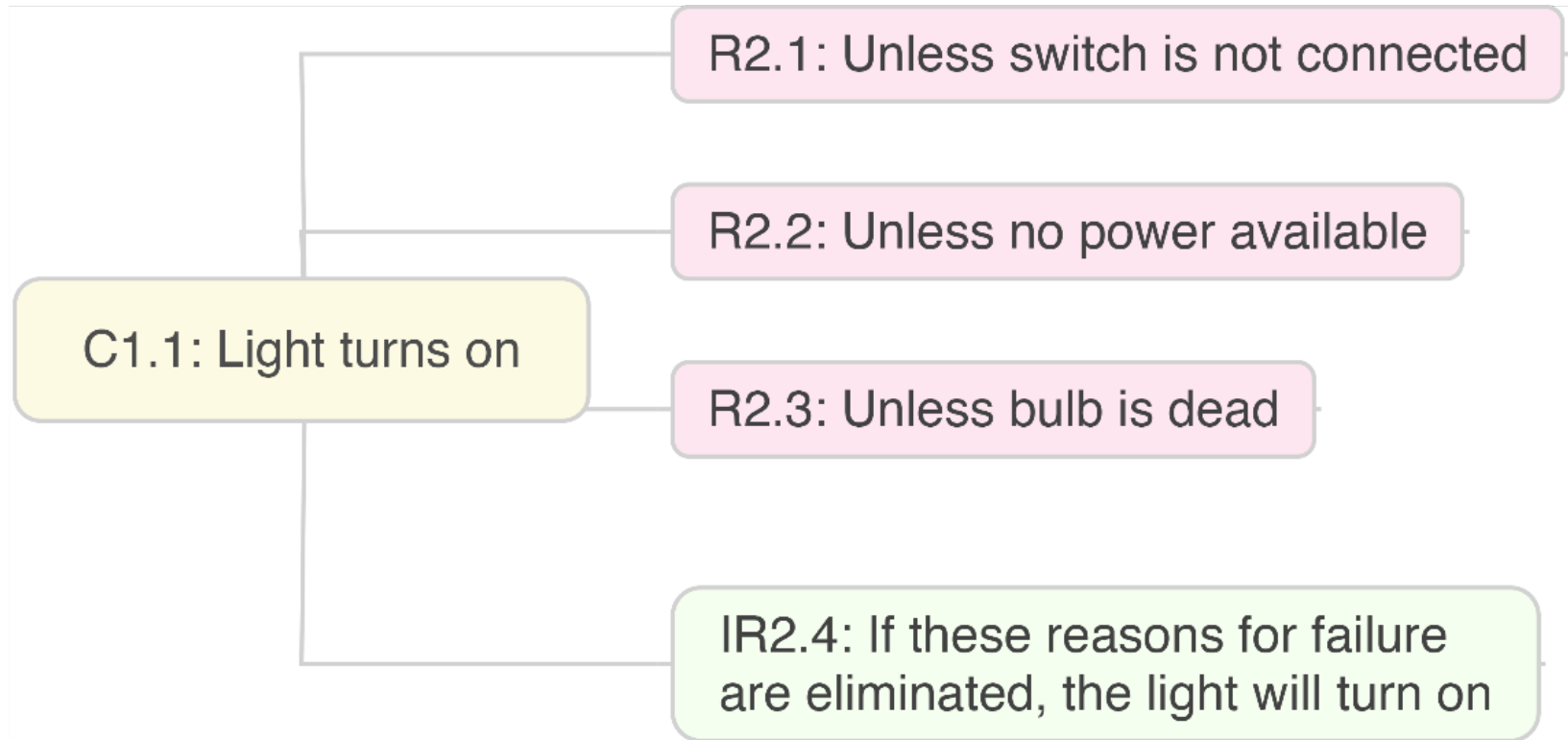
An eliminative argument is visualized in a *confidence map*, which shows reasons for doubt graphically.

It's a reframing of an assurance argument to make doubts explicit.

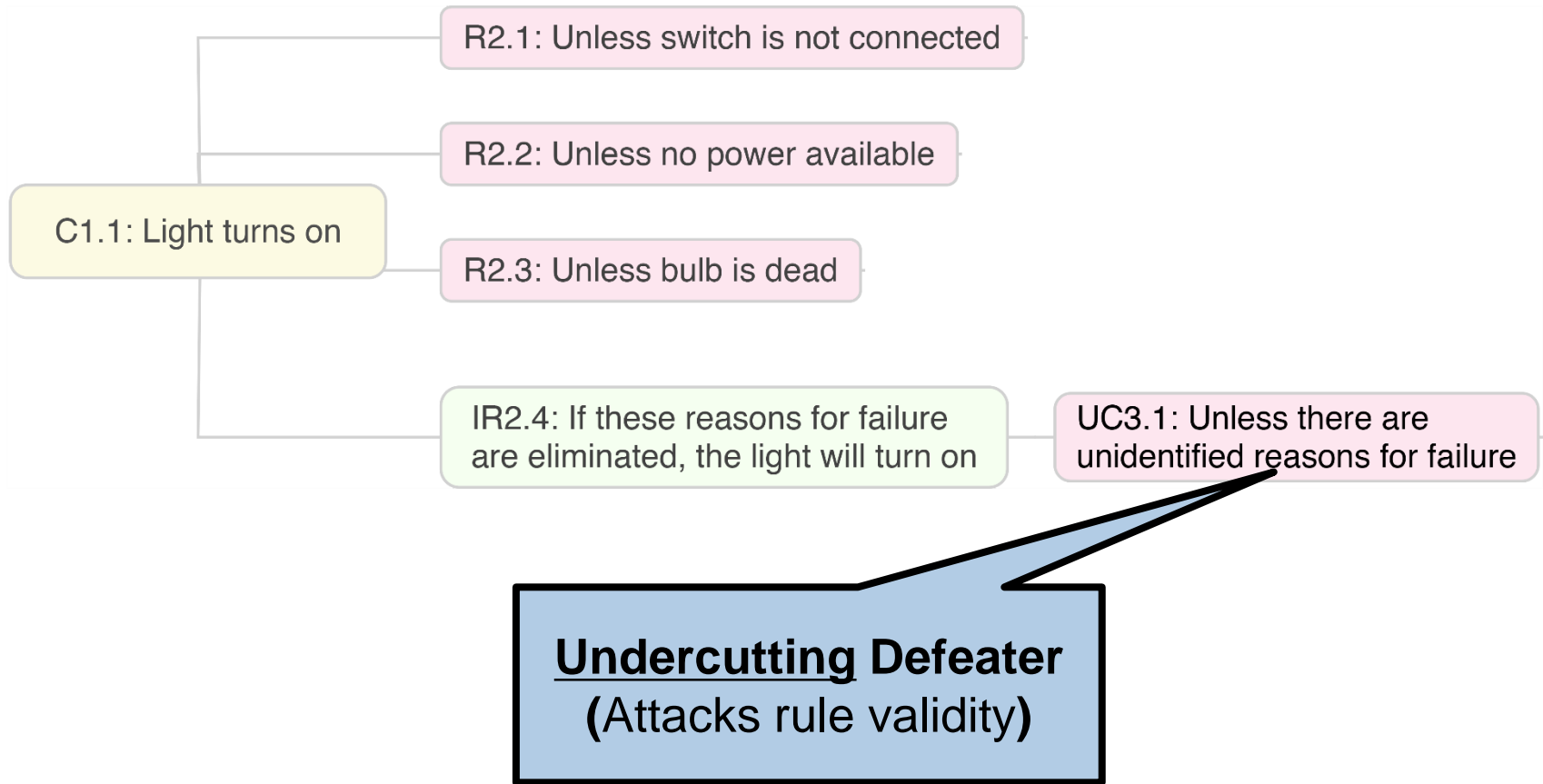
# A Confidence Map



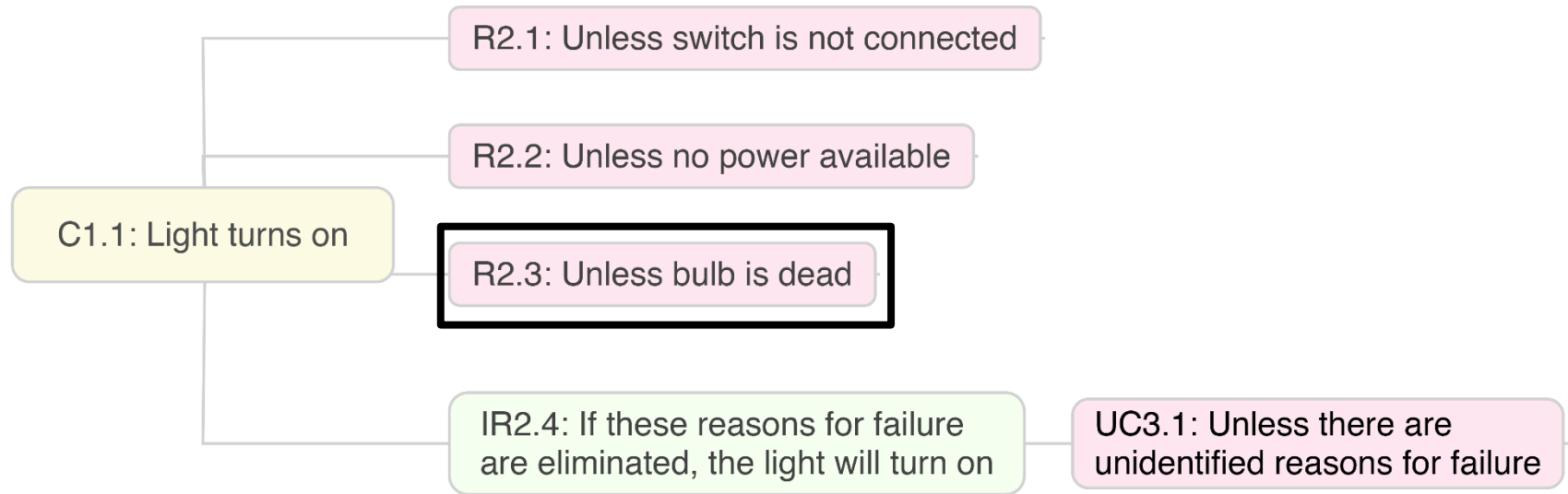




**Inference Rule**  
**(makes reasoning explicit)**



**Cuts part of the *connection* between a premise and a conclusion**



R2.3: Unless bulb is dead

Ev3.1: Examination results showing bulb doesn't rattle when shaken

UM4.1: But the examiner is hard of hearing

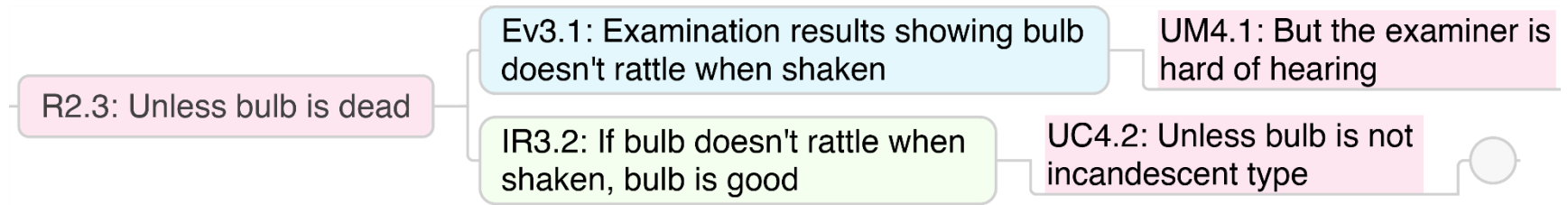
R2.3: Unless bulb is dead

Ev3.1: Examination results showing bulb doesn't rattle when shaken

UM4.1: But the examiner is hard of hearing

**Undermining Defeater**  
**(Attacks evidence validity)**

**Evidence is the foundation of an argument; undermining defeaters weaken this foundation**



1|2

R2.3: Unless bulb is dead

0|1

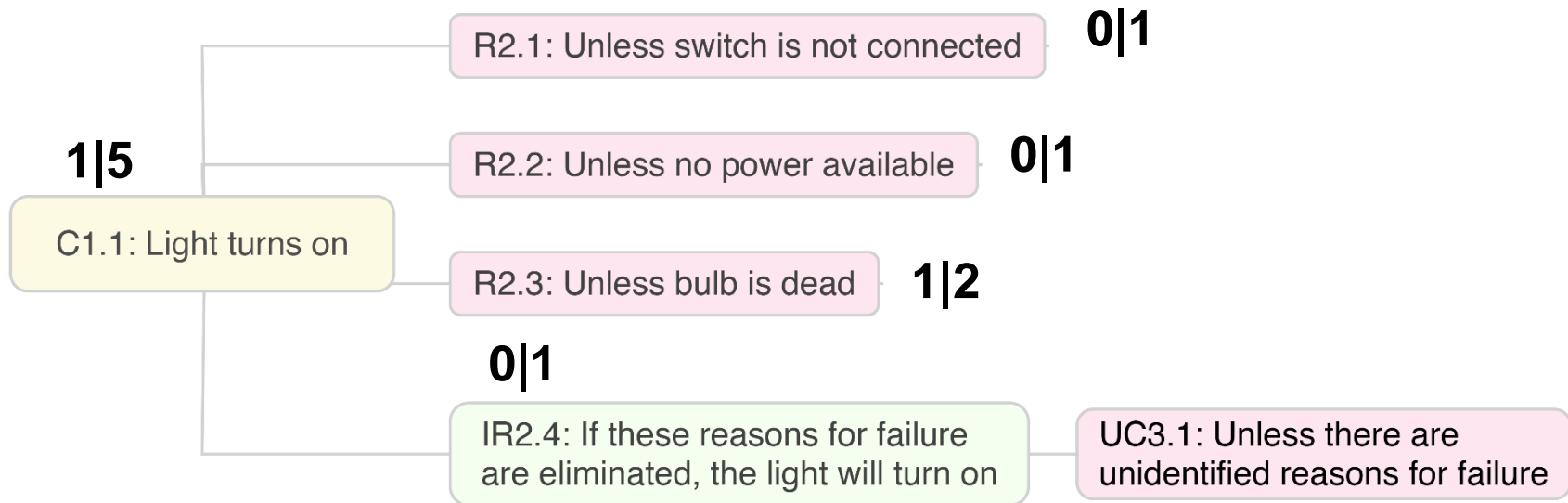
Ev3.1: Examination results showing bulb doesn't rattle when shaken

UM4.1: But the examiner is hard of hearing

IR3.2: If bulb doesn't rattle when shaken, bulb is good

UC4.2: Unless bulb is not incandescent type

1|1



# Sources of Doubt (Defeasible Reasoning)

## Finding doubts

- Attack claim (rebutting defeater) — why claim may be **false**
- Attack evidence (undermining defeater) — why evidence may be **compromised**
- Attack inference (undercutting defeater) — premise ok; **conclusion uncertain**

Confidence increases as **reasons for doubt** (defeaters) are eliminated

- $0/n$  (no confidence) — **no reasons for doubt eliminated**
- $n/n$  (complete confidence) — **no doubts remain**
- $2/3$  (partial confidence) — ***residual doubt***

# So What?

Evidence and argument is used to gain confidence that a system has a desirable property, **BUT**

- What constitutes “good” or “strong” evidence?
- What is a strong argument? Where are its weaknesses?
- What is a justifiable basis for having confidence?

Evidence is valuable only to the extent it eliminates

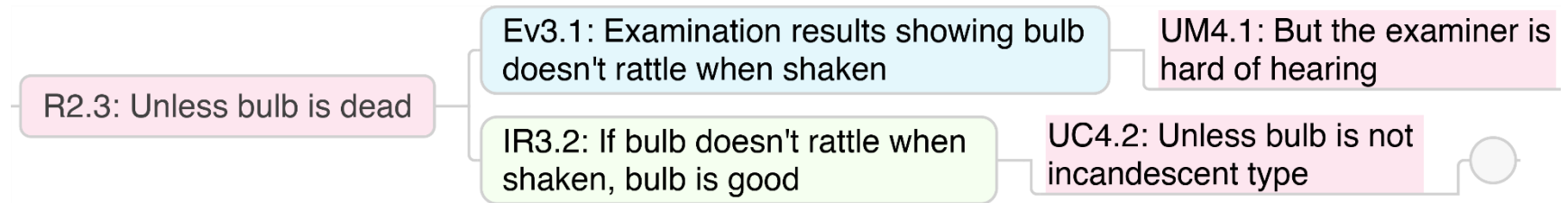
- doubts about the **validity of claims**
  - rebutting defeaters
- doubts about **argument soundness** (weaknesses in evidence and inference rules)
  - undercutting and undermining defeaters

# A More Nuanced Evaluation of Defeaters

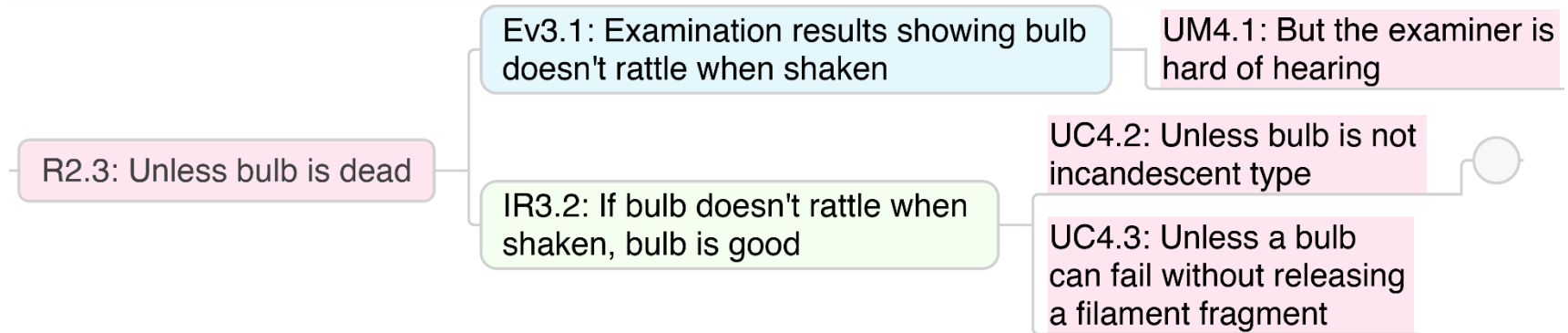
Some defeaters are more likely or more important than others

How could this be handled?

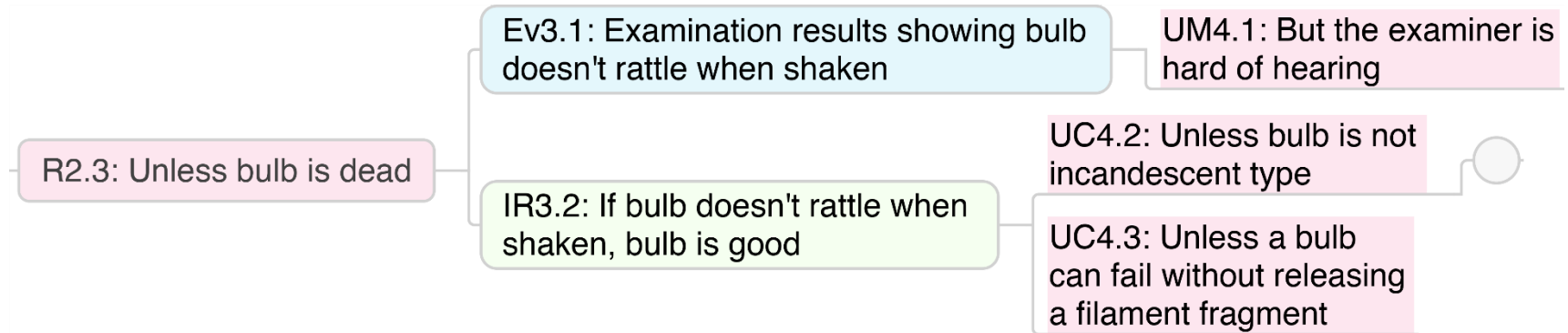
# Partial Defeater Elimination



# Partial Defeater Elimination

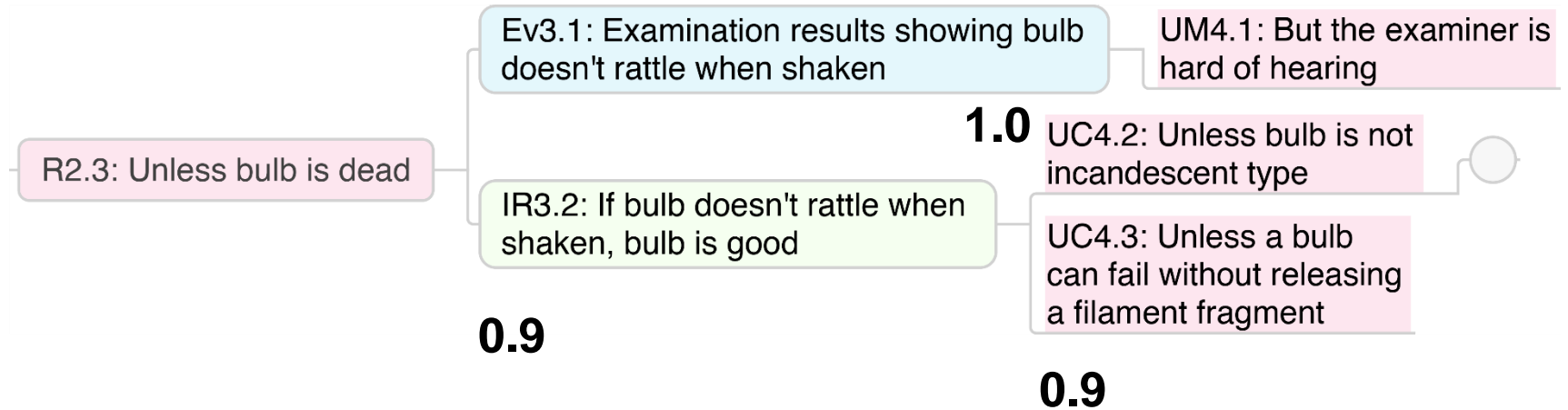


# Partial Defeater Elimination

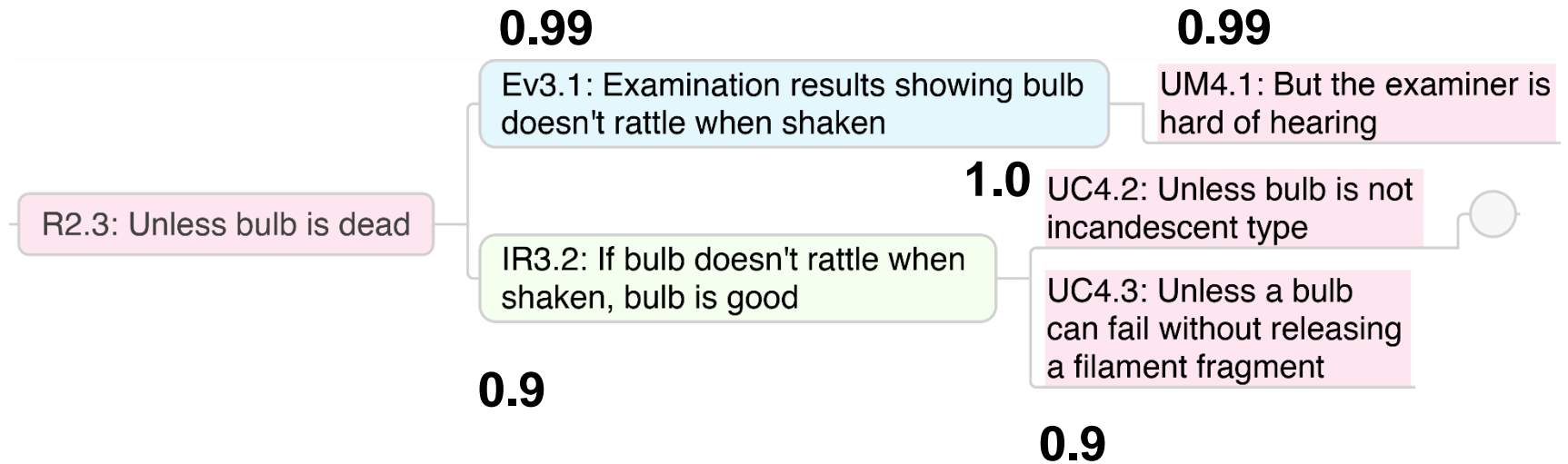


**0.9**

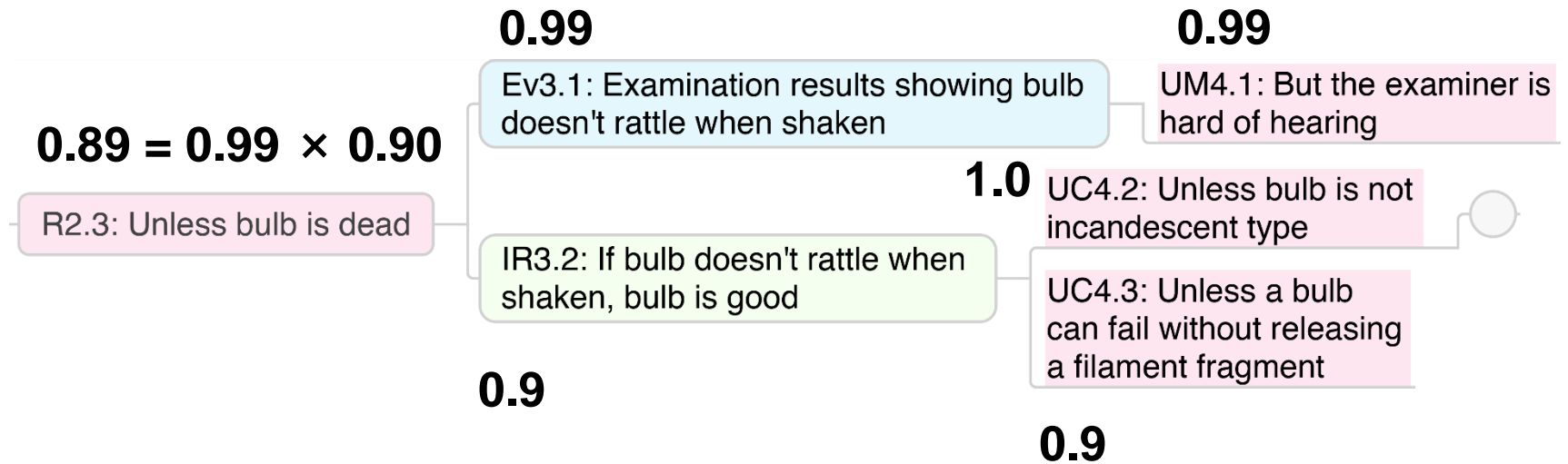
# Partial Defeater Elimination



# Partial Defeater Elimination



# Partial Defeater Elimination



# Defeater Elimination

**With eliminative argumentation, we are explicit about:**

- **how a plan can fail**
- **potential flaws in an argument showing why failure is unlikely**

**And we know what evidence we need to eliminate**

- **sources of failure (rebutting defeaters) and**
- **sources of argument weakness (evidence and inferences)**

# What if a Evidence Validates a Defeater?

Example: a test fails (counterevidence)

Potential response

- Eliminate the counterevidence
  - Fix the system (e.g., prior to deployment)
  - Restrict a contradicted claim (e.g., don't use the system under certain conditions)
  - Undermine the counterevidence (the system is actually working; the reported error is not an error)
- Accept the counterevidence
  - Live with reduced confidence by modifying the argument
    - Add a new defeater and/or inference rule
    - “Uneliminate” an existing defeater (decrease its probability of elim.)
  - Don't change the argument
    - The counterevidence is consistent with an uneliminated defeater
    - An *uneliminated* defeater has the same effect on confidence as a *true* defeater

# Gaining Assurance After a System Change

Can analyze undermining defeaters to see what evidence may no longer be valid after the change

- Retest the hearing of the examiner after a bulb change
- Do a shake test for a new bulb to ensure that the bulb that is now in the system is the bulb that was tested

Can analyze undercutting defeaters to see which inference rules are no longer valid

- If an LED bulb is used, lack of a rattle is no longer relevant evidence
- May need to add new rebutting defeaters when inference rule is changed (e.g., new failure modes have been introduced)

# Objections

What if a relevant defeater has not been identified?

What if a defeater cannot be completely eliminated?

Not all defeaters are of equal importance. How is this handled?

Just counting defeaters ( $m/n$ ) seems weak (compared to Bayesian probability or enumerative induction). What is being gained (and lost) with this approach?

The potential number of defeaters seems incredibly large for a real system. Is this approach practical?

# Objections

## What if a relevant defeater has not been identified?

What if a defeater cannot be completely eliminated?

Not all defeaters are of equal importance. How is this handled?

Just counting defeaters ( $m/n$ ) seems weak (compared to Bayesian probability or enumerative induction). What is being gained (and lost) with this approach?

The potential number of defeaters seems incredibly large for a real system. Is this approach practical?

# What if there is an unidentified defeater?

Confidence cases are inherently defeasible; always possible something has been omitted

- Complete confidence ( $n|n$ ) only reflects what is known at a particular point in time

Uncertainty about completeness is itself a reason for doubt that needs to be recognized and countered

- “Not all hazards have been identified”
- Assessment of a case must consider this as a reason for doubting the adequacy of the case
  - Eliminative argumentation provides a method for identifying where sources of doubt can be found

Eliminative argumentation provides ways of thinking about and explaining why one should have confidence in a case, or a claim

- The approach does not, of course, guarantee a sound case
- But helps in developing sound and complete arguments

# Objections

What if a relevant defeater has not been identified?

**What if a defeater cannot be completely eliminated?**

Not all defeaters are of equal importance. How is this handled?

Just counting defeaters ( $m/n$ ) seems weak (compared to Bayesian probability or enumerative induction). What is being gained (and lost) with this approach?

The potential number of defeaters seems incredibly large for a real system. Is this approach practical?

# Incomplete Defeater Elimination

We have addressed this in our examples

We accept that in practical cases, there will always be some residual doubt

- The issue is whether the remaining doubts are considered significant or not

The general principle is that uneliminated lower level doubts propagate to higher level claims

- The goal is to formulate lower level defeaters that *can* be eliminated by appropriate evidence and inference rules

# Objections

What if a relevant defeater has not been identified?

What if a defeater cannot be completely eliminated?

**Not all defeaters are of equal importance. How is this handled?**

Just counting defeaters ( $m/n$ ) seems weak (compared to Bayesian probability or enumerative induction). What is being gained (and lost) with this approach?

The potential number of defeaters seems incredibly large for a real system. Is this approach practical?

# Differential Defeater Importance

The elimination of some defeaters seems more important (in some intuitive sense) than others. A strict eliminative induction approach treats all uneliminated defeaters equally.

- Consider hazards identified in a safety analysis. All above a certain threshold must be eliminated/mitigated
  - Assessing their relative importance/likelihood is not profitable

# Objections

What if a relevant defeater has not been identified?

What if a defeater cannot be completely eliminated?

Not all defeaters are of equal importance. How is this handled?

**Just counting defeaters ( $m/n$ ) seems weak. What is being gained (and lost) with this approach?**

The potential number of defeaters seems incredibly large for a real system. Is this approach practical?

# Why Use Eliminative Argumentation

With eliminative argumentation, we learn something concrete about why a system works, namely, eliminated defeaters

- With enumerative induction (e.g., operational experience), we only learn something statistical. (A different kind of useful knowledge)

Eliminative argumentation avoids “confirmation bias”

- To the extent evidence eliminates defeaters, we know an argument is valid
  - For at least the situations covered by these defeaters
- But our confidence is limited by the uneliminated defeaters
  - So, just because some things work adequately, the uneliminated defeaters constrain us from overgeneralizing the implications of success, e.g., successful test executions

# Objections

What if a relevant defeater has not been identified?

What if a defeater cannot be completely eliminated?

Not all defeaters are of equal importance. How is this handled?

Just counting defeaters ( $m/n$ ) seems weak (compared to Bayesian probability or enumerative induction). What is being gained (and lost) with this approach?

**The potential number of defeaters seems incredibly large for a real system. Is this approach practical?**

# Practical Considerations

The amount of evidence and argument for a real system is inherently quite large

An eliminative argument provides an overview and reviewable summary of why someone should have confidence in your approach

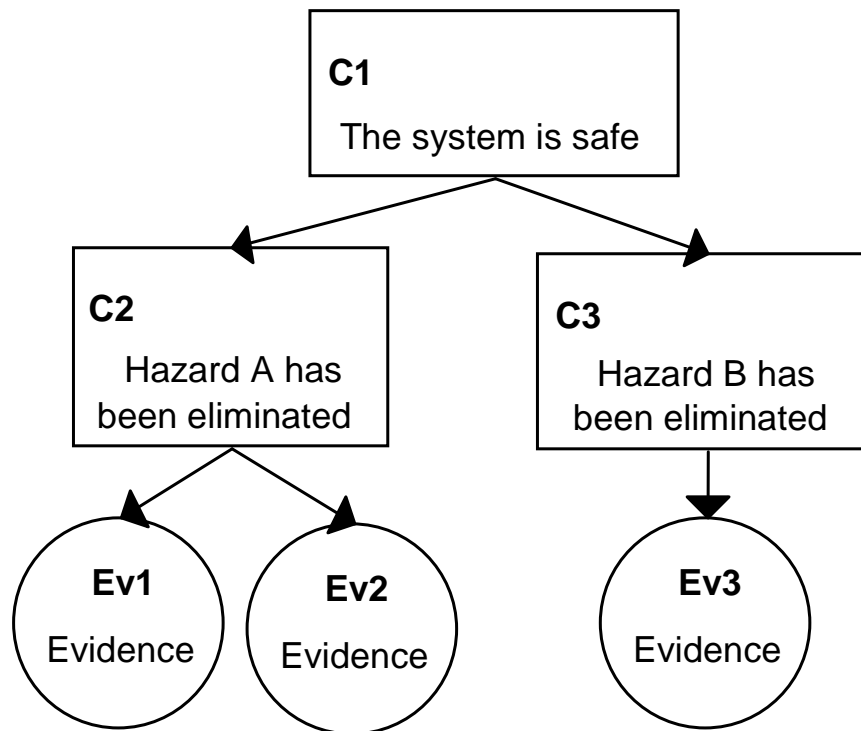
# REVIEW

# The Problem

How confident in C1? Why? (Number of uneliminated doubts)

What does it mean to have confidence? (To have no doubt)

What could be done to improve confidence? Why? (Elim. more doubts)



# Sources of Doubt (Defeasible Reasoning)

## Finding doubts

- Attack claim (rebutting defeater) — why claim may be **invalid**
- Attack evidence (undermining defeater) — why evidence may be **compromised**
- Attack inference (undercutting defeater) — premise ok; **conclusion uncertain**

Confidence increases as **reasons for doubt** (defeaters) are eliminated

- $0/n$  (no confidence) — **no reasons for doubt eliminated**
- $n/n$  (complete confidence) — **no doubts remain**
- $2/3$  (partial confidence) — ***residual doubt***

# Summary

Eliminative argumentation (identification of doubts and their elimination)

- Provides a framework for building confidence in properties of a system
- Confidence maps are a visualization of an eliminative argument
- Confidence maps explicitly document reasons for doubt and their elimination

**Confidence maps provide justification for  
belief in a claim**

We do so by identifying and eliminating defeaters (doubts)  
relevant to the claim and the argument

# CREATING AND REVIEWING

# Creating Good Confidence Maps

## Checklists for well-formed

- Claims
- Evidence
- Defeaters of the various types
  - Rebutting (applicable to claims)
  - Undercutting (applicable to inference rules)
  - Undermining (applicable to evidence)

# CM Guidance: Claims

Claim is stated as a predicate, usually in the present tense

- States what you wish to be true
- Should be a simple statement, with explanation of terms in a *context* element
  - Example: “System is acceptably reliable”
  - Context clause defines “acceptably reliable” and may refer to various reports and other documents

# CM Guidance: Rebutting Defeaters

Rebutting defeaters (always associated with a claim)

- Phrased as a predicate introduced by “Unless”
  - Example: “Unless bulb is dead”
- Should express *a reason* for deciding that a claim is invalid
  - The claim is valid to the extent we have confidence that the defeater is false (i.e., eliminated)
  - Should not merely be the opposite of the claim (this adds no useful information)

# CM Guidance: Inference Rules

Given claim, C, inference rules have the form

**“if  $defeater_1$  and  $defeater_2$  and ... then C”**

where  $defeater_i$  are rebutting defeaters associated with claim C

The inference rules are implied by the structure of a CM, but it can be helpful to state them explicitly:

- If they capture a generalization (e.g., “if all failure modes are eliminated, the system is reliable” rather than repeating all the failure modes)
- Explicit statement can be helpful while developing an argument
- They serve as a place to attach reasons why the inference rule might not always imply its conclusion (i.e., undercutting defeaters)

# CM Guidance: Undercutting Defeaters

Undercutting (UC) defeaters (always associated with a inference rule)

- Phrased as a predicate introduced by “Unless”
  - Example rule: “If all identified reasons for failure are eliminated, the system is reliable”
  - Example UC defeater: “Unless there are unidentified reasons for failure”
- If the defeater is not eliminated, we **don’t know** whether the rule’s conclusion is valid
  - The rule is valid to the extent we have confidence that the defeater is eliminated
  - If the defeater is true, *we don’t know whether the conclusion holds*
    - Example: An unidentified reason for failure doesn’t necessarily mean a system is unreliable (the failure might never occur)

# CM Guidance: Evidence

Evidence is expressed in the form

[Noun phrase] showing P

Where the “Noun phrase” describes the data comprising the evidence

- E.g., “Examination results”

P is a predicate asserting an interpretation of the data relevant to the argument

- E.g., “bulb doesn’t rattle when shaken”

The noun phrase serves only to identify information whose interpretation (P) is relevant to an argument

# CM Guidance: Undermining Defeaters

Undermining defeaters give reasons for doubting the validity or significance of the evidence and/or its interpretation

The reasons for doubt are predicates, preceded by the word “But”

- Example: “But the examiner is hard of hearing”
  - The evidence data: Examination results (i.e., examiner’s report)
  - The evidence claim: “Bulb doesn’t rattle when shaken”

# Terminating an Argument

At some point, must assume some defeater is eliminated

- By documenting this assumption, the argument can later be challenged or the effect of a changed assumption can be traced
  - Example: Only incandescent bulbs are being used

# Contact Information

## John B. Goodenough

Fellow

Telephone: +1 412-390-4043

Email: [jbg@sei.cmu.edu](mailto:jbg@sei.cmu.edu)

## U.S. Mail

Software Engineering Institute

4500 Fifth Avenue

Pittsburgh, PA 15213-2612

USA

## Charles B. Weinstock

Principal Researcher

Telephone: +1 412-268-7719

Email: [weinstock@sei.cmu.edu](mailto:weinstock@sei.cmu.edu)