2nd International Workshop on Argument for Agreement and Assurance (AAA 2015), Kanagawa Japan, November 2015

On the Interpretation Of Assurance Case Arguments

John Rushby

Computer Science Laboratory
SRI International
Menlo Park, CA

Introduction

- I'm focused on the assurance and certification of software for commercial airplanes
- Currently assured by DO-178C
  - Enumerates 71 "objectives" that must be satisfied for the most critical software
  - e.g., "Ensure that each High Level Requirement (HLR) is accurate, unambiguous, and sufficiently detailed, and that the requirements do not conflict with each other" [Section 6.3.1.b]
- It seems to work: no incidents due to flaws in software implementation
  - DO-178C is about correctness of implementation wrt HLR
  - ARP 4754 and others are concerned with safety of HLR

Introduction (ctd.)

- But the world is changing
  - NextGen integrates once separate air and ground systems
  - Unmanned vehicles in same airspace
  - More autonomous systems
  - New methods of software development and assurance
- We don't really know why DO-178C works
  - So difficult to predict impact of changed environment
  - And difficult to update (10 years to go from B to C)
- So look at Assurance Cases as a possible way forward
  - Retrospective reformulation of DO-178C as an assurance case (Michael Holloway)
  - Then look for a scientific basis to assurance cases

Assurance Cases

- The idea is that we "make the case" to justify deployment of some system by
  - Stating the claim that it must satisfy
    - ⋆ Generally safety- or correctness-related
  - Developing evidence about its assumptions, design, implementation, performance etc.
  - Constructing a structured argument that justifies the claim, based on the evidence
- How should we interpret these arguments?
- And what are the expectations on them?
  - "compelling, comprehensible and valid" [00-56]
  - Are these all the same?

Complications: Inductive and Deductive Arguments

- The world is an uncertain place (random faults and events)
- Our knowledge of the world is incomplete, may be flawed
- Our reasoning may be flawed also
- So an assurance case cannot expect to prove its claim
- Hence, the overall argument is inductive
  - Evidence & subclaims strongly suggest truth of top claim
- Rather than deductive
  - Evidence & subclaims imply or entail the top claim

Complications: Confidence Items

- If the overall argument is inductive
- Does that mean all its steps may be inductive too?
- Traditionally, yes!
  - Considered unrealistic to be completely certain
  - cf. ceteris paribus hedges in science
- Can add ancillary confidence items to bolster confidence in inductive steps
  - Evidence or subclaims that do not directly contribute to the argument
  - i.e., their falsity would not invalidate the argument
  - But their truth increase our confidence in it
- Eh?

Complications: Graduated Assurance

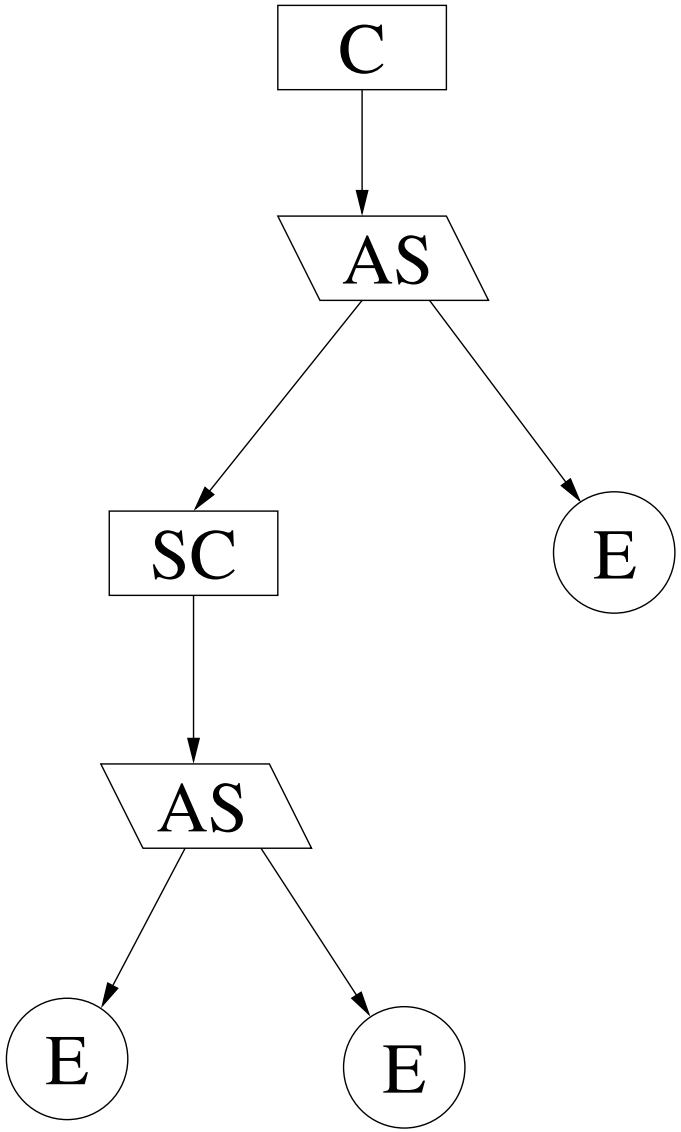
- Assurance is expensive, so most standards and guidelines allow less assurance effort for elements that pose lesser risks
- E.g. DO-178C
  - 71 objectives for Level A, 33 with independence
  - 69 objectives for Level B, 21 with independence
  - 62 objectives for Level C, 8 with independence
  - 26 objectives for Level D, 5 with independence
- So if Level A is "compelling, comprehensible and valid"
- The lower levels must be less so, or not so
- We need some idea what is lost, and a measure of how much

Proposed Interpretation

- Clearly need a semantics to account for all this
- I'm going to propose a simple, even obvious, semantics for a sound assurance case
- I further propose that only sound assurance cases should be accepted
- However, sound assurance cases can have different strengths

Structured Argument

In a generic notation (GSN shapes, CAE arrows)



C: Claim

AS: Argument Step

SC: Subclaim

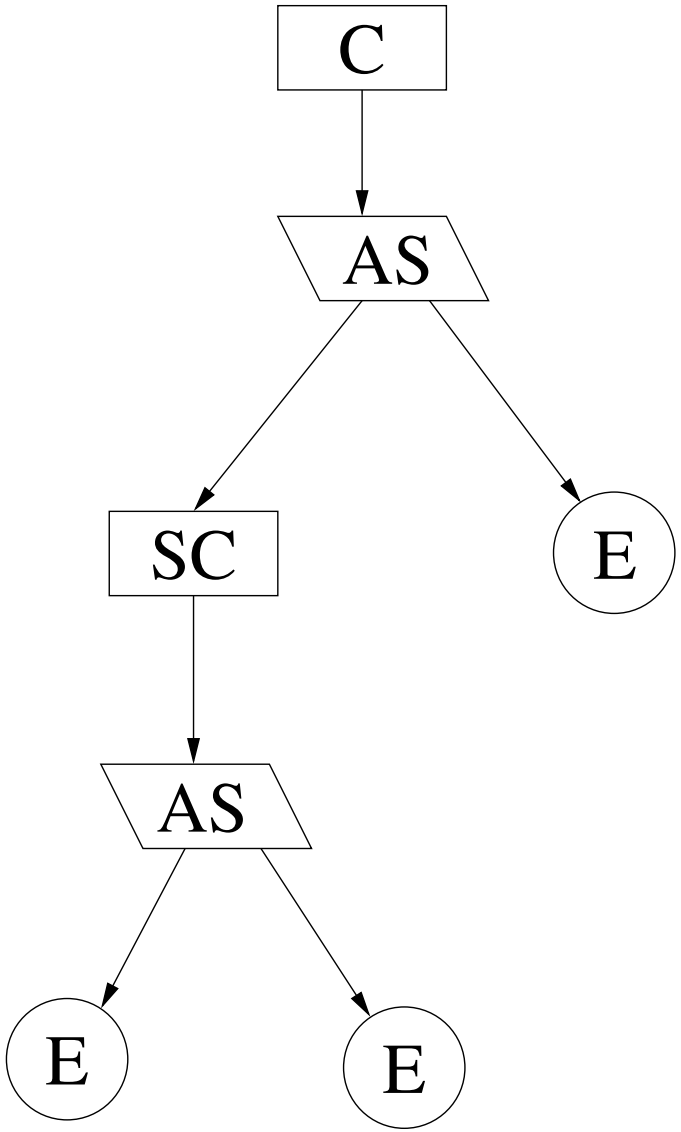
E: Evidence

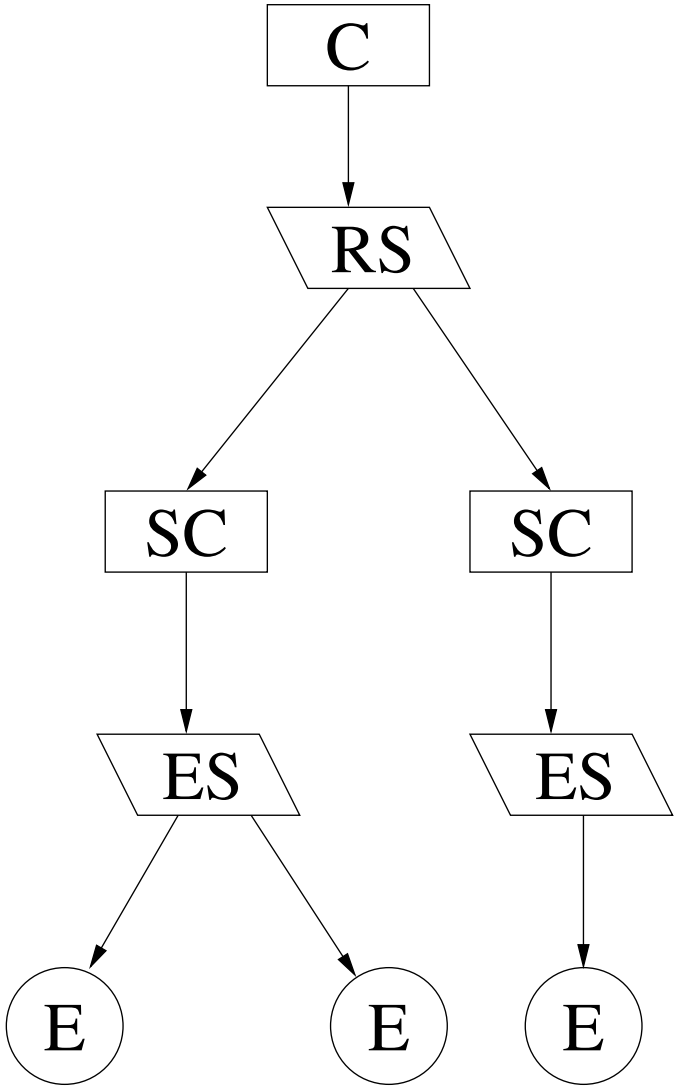
A hierarchical arrangement of argument steps, each of which justifies a claim or subclaim on the basis of further subclaims or evidence

Argument Steps and Layered Arguments

- We decompose top-level claim into conjunction of subclaims
- And iterate
- Until we get down to subclaims supported by evidence
- Provide a narrative justification for each step
- Easier to understand when just two kinds of argument steps
  - Reasoning steps: subclaim supported by further subclaims
  - Evidential steps: subclaim supported by evidence
- Call this a simple form argument
  - Can normalize to this form by adding subclaims
  - In the paper I explain how to give a direct interpretation

Normalizing an Argument to Simple Form





RS: reasoning step; **ES**: evidential step

Why Focus on Simple Form?

- The two kinds of argument step are interpreted differently
- Evidential steps
  - These are about epistemology: knowledge of the world
  - Bridge from the real world to the world of our concepts
  - Have to be considered inductive
  - Multiple items of evidence are "weighed" not conjoined
- Reasoning Steps
  - These are about logic/reasoning
  - Conjunction of subclaims leads us to conclude the claim
    - ⋆ Deductively: subclaims imply claim (my preference)
    - ⋆ Inductively: subclaims suggest claim
- Combine these to yield complete arguments
  - Those evidential steps whose weight crosses some threshold of credibility are treated as premises in a classical deductive interpretation of the reasoning steps

Weighing Evidential Steps

- We measure and observe what we can
  - e.g., test results
- To infer a subclaim that is not directly observable
  - e.g., correctness
- Different observations provide different views
  - Some more significant than others
  - And not all independent
- "Confidence" items can be observations that vouch for others
  - Or provide independent backup
- Need to "weigh" all these in some way
- Probabilities provide a convenient metric
- And Bayesian methods and BBNs provide tools

The Weight of Evidence?

- Plausible to suppose that we should accept claim C given evidence E when $P(C \mid E)$ exceeds some threshold
- These are subjective probabilities expressing human judgement
- Experts find $P(C \mid E)$ hard to assess
- And it is influenced by prior $P(C)$, which can express ignorance. . . or prejudice
- Instead, factor problem into alternative quantities that are easier to assess and of separate significance
- So look instead at $P(E \mid C)$
  - Related to $P(C \mid E)$ by Bayes' Rule
  - But easier to assess likelihood of observations given claim about the world than vice versa

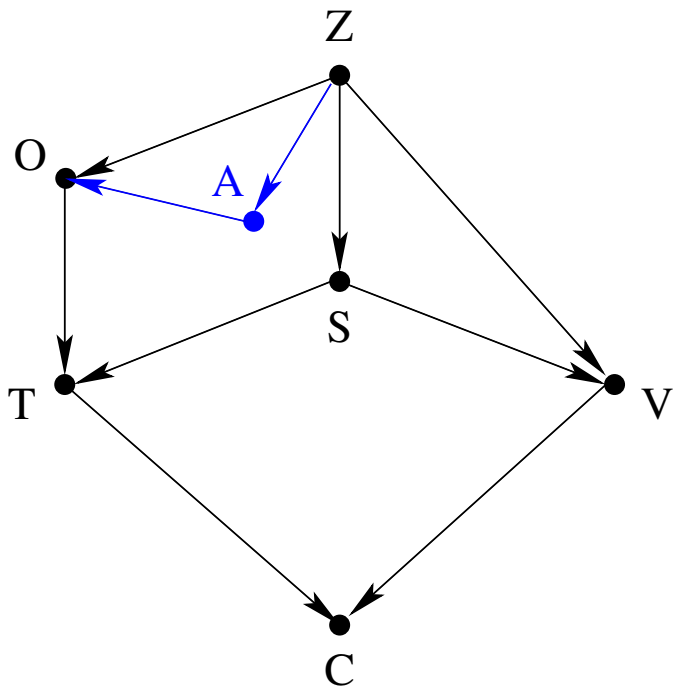
Confirmation Measures

- We really are interested in the extent to which E supports C . . . rather than its negation $\neg C$
- So focus on the ratio or difference of $P(E \mid C)$ and $P(E \mid \neg C)$, . . . or logarithms of these
- These are called confirmation measures
- They weigh C and $\neg\, C$ "in the balance" provided by E
- Suggested that these are what criminal juries should be instructed to assess (Gardner-Medwin)
- Good's measure: $\log \dfrac{P(E \mid C)}{P(E \mid \neg\, C)}$
- Kemeny and Oppenheim's measure: $\dfrac{P(E \mid C) - P(E \mid \neg\, C)}{P(E \mid C) + P(E \mid \neg\, C)}$
- Much discussion on merits of these and other measures

Application of Confirmation Measures

- I do not think the specific measures are important
- Nor do I advocate applying these methods to the evaluation of individual arguments
- Rather, use BBNs and confirmation measures for what-if investigations
  - Can help in selection of evidence for evidential steps
  - e.g., refine what objectives DO-178C should require
- Example (next slides) use of "artifact quality" objectives as confidence items in DO-178C

Weighing Evidential Steps With BBNs



Z: System Specification

O: Test Oracle

S: System's true quality

T: Test results

V: Verification outcome

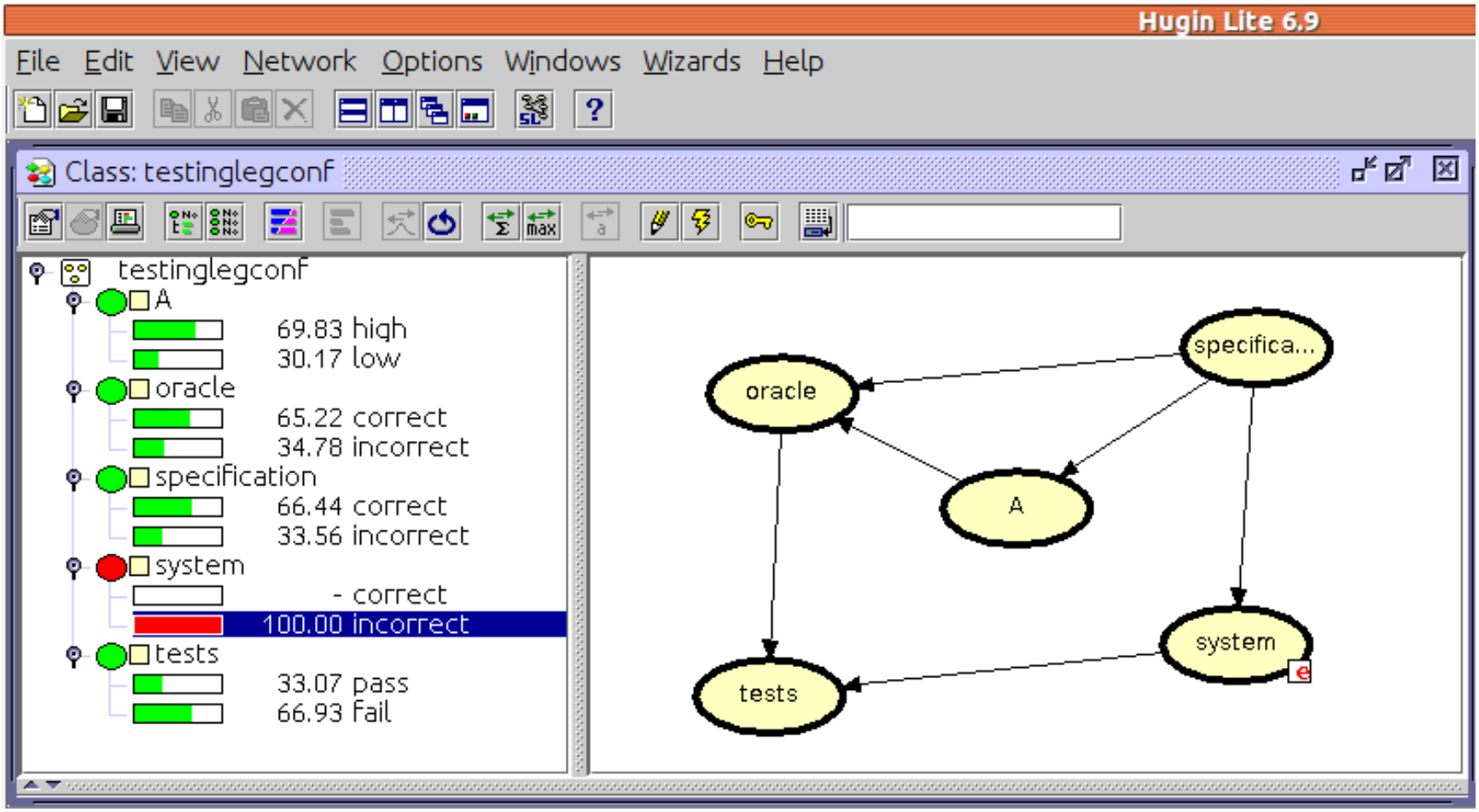
A: Specification "quality"

C: Conclusion

Example joint probability table: successful test outcome

Correct System		Incorrect System	
Correct Oracle	Bad Oracle	Correct Oracle	Bad Oracle
100%	50%	5%	30%

Example Represented in Hugin BBN Tool



Evaluating Reasoning Steps

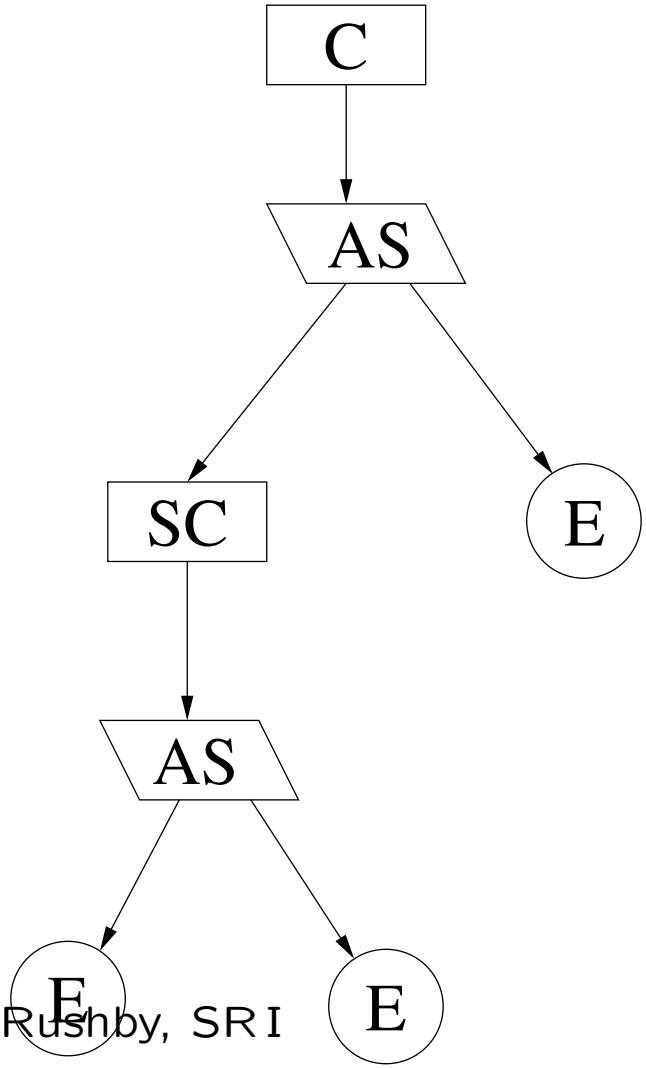
- When all evidential steps cross our threshold for credibility, we use them as premises in a classical interpretation of the reasoning steps
  - Deductive: p_1 AND p_2 AND $\cdots$ AND p_n IMPLIES c
  - Inductive: p_1 AND p_2 AND $\cdots$ AND p_n SUGGESTS c
- I advocate the deductive interpretation, for three reasons
  - There is no classical interpretation for inductive reasoning
    - ⋆ Many proposals: Dempster-Shafer, fuzzy logic, probability logic
    - ⋆ But none universally accepted
    - ⋆ And they flatten the argument (forthcoming slide)
  - Inductive reasoning is not modular: must believe either the gap is insignificant (so deductive), or taken care of elsewhere (so not modular)
  - There is no way to evaluate the size of the gap in inductive steps (next slide)

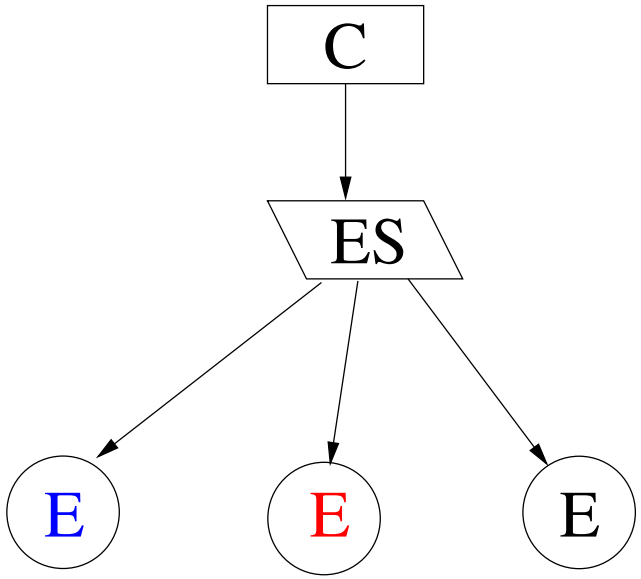
The Inductive Gap

- Must surely believe inductive step is nearly deductive and would become so if some missing subclaim or assumption a were added
  - p_1 AND p_2 AND $\cdots$ AND p_n SUGGESTS c
  - a AND p'_1 AND p'_2 AND $\cdots$ AND p'_n IMPLIES c
- If we knew anything at all about a it would be irresponsible not to add it to the argument
- Since we did not do so, we must be ignorant of a
- Follows that we cannot estimate the doubt in inductive argument steps

Probabilistic, Fuzzy and D-S Interpretations

- Insensitive to logical content of reasoning steps
- Effectively replace each subclaim by its supporting evidence
- Thereby flattening the argument





Flattened Arguments

- There's a reason we don't do this
  - An assurance case is not just a pile of evidence
    - ⋆ That's DO-178C, for example
  - It is an argument
  - With a structure based on our reasoning about the system

- So the reasoning steps should be interpreted in logic

Graduated Assurance

- I'll say an assurance case is valid if its reasoning steps are judged to be deductively valid
  - Expect to see justification in some form
- A valid case is sound if in addition its evidential steps cross the threshold for credibility
  - All inductive doubts located here
- For graduated assurance, need some additional notion of argument strength
- One approach to weakening an argument for lower levels is to reduce the threshold on evidential steps
- But others actually change the argument
  - E.g., Level D of DO-178C removes the Low Level Requirements (LLR) and all attendant steps

Evaluating Argument Strength Under Reduced Thresholds

- Although I don't advocate flattening then BBNs
  - As a way to evaluate soundness of an argument
- It could be a way to quantify strength of a sound argument
- More simply
  - Just sum (Adams' Uncertainty Accumulation)
  - Or multiply (independence assumption)

  The probabilities calculated (by BBNs) for evidential steps
- Beware of gaming:
  - Combining subclaims to maximize strength measure
- Could do this on an ordinal scale (low, medium, high, etc.)
- Note that it's a weakest link calculation
- Graduated assurance retains soundness, reduces strength

Evaluating Argument Strength Under Changes

- Recall Level D of DO-1788C changes the argument
  - Removes everything to do with LLR
- Reason for LLR is not just more evidence, but the credibility of the overall argument strategy
  - More credible to go from HLR to EOC via LLR (Levels A, B, C)
  - Than in a single leap (Level D)
- So there's more to it than just evidential strength
  - Topic for future work: related to ability to withstand defeaters

Conclusion

- Interpretation is a combination of probability and logic
- (Possibly informal) probabilities for evidential steps
- Logic for reasoning steps
- Case is sound **if** evidential steps cross some threshold **and** reasoning steps are deductively valid
  - All inductive doubt is located in the evidential steps
  - Inductive reasoning steps are too low a bar
- Graduated Assurance may weaken evidential support
  - Overall strength of a sound case is then determined by weakest evidential step
  - Can formalize this in probability logic, but I think the real appeal has to be to intuition and consensus. . .
- Deeper notion of strength needed for other forms of graduated assurance: defeaters and argumentation frameworks may be the way to go here

Links

- Lengthy report: http://www.csl.sri.com/~rushby/abstracts/assurance-cases15

- What do you think?