

Cognitive Bias in Systems Verification

Steve Larson

Jet Propulsion Laboratory
California Institute of Technology

March 5, 2012

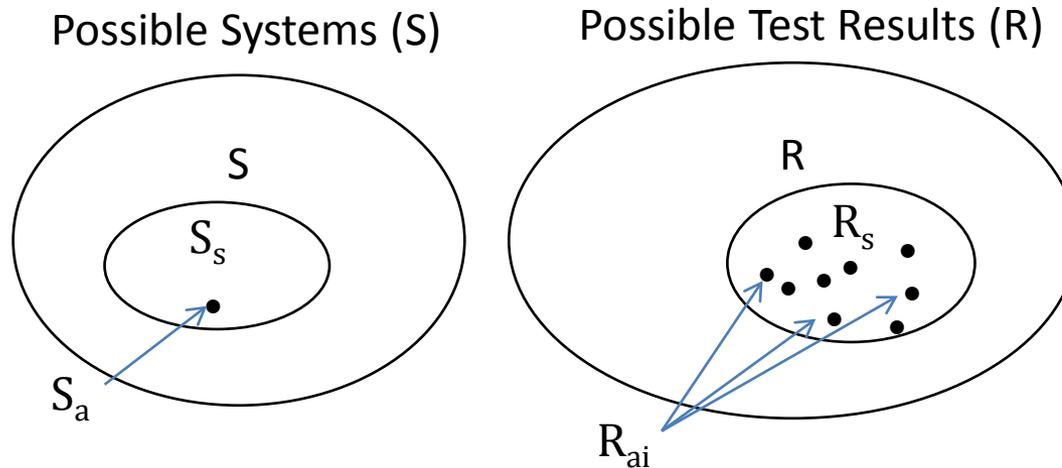
Introduction

- Background
- Formulation of the verification problem
- Discussion and examples of selected biases
 - Probability and representativeness
 - Positive Test Strategies
 - Availability
 - Overconfidence
 - Anchoring and Adjustment
- Debiasing
- Conclusion

Background

- Working definition of cognitive bias
 - Patterns by which information is sought and interpreted that can lead to systematic errors in decisions
- Cognitive bias is used in diverse fields
 - Economics, Politics, Intelligence, Marketing, to name a few...
- Attempts to ground cognitive science in physical characteristics of the cognitive apparatus exceed our knowledge
 - Studies based on correlations; strict cause and effect is difficult to pinpoint
 - Effects cited in the paper and discussed here have been replicated many times over, and appear sound
 - Many biases have been described, but it is still unclear whether they are all distinct
 - There may only be a handful of “fundamental” biases, which manifest in various ways
- Existence of bias most often explained in terms of evolutionary advantage
 - Efficiency
 - “Good enough”
 - Enhanced likelihood of making important advances
- Bias is never the whole story

The Verification Problem

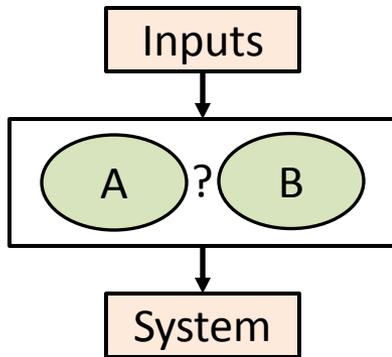


- Given a set of test results R_{ai} that belong to the set of satisfactory test results R_s , What is the probability that the actual system S_a belongs to the set of satisfactory systems S_s ?

Probability and Representativeness

- Representativeness is the phenomenon of drawing conclusions based on similarities between available data and other beliefs or knowledge relating to the outcome
 - Example 1: the coin toss sequence THHTHT is often judged to be more likely than TTTTHH because it appears more random
 - Example 2: Estimates of the likelihood that an individual is an engineer rather than a lawyer change significantly when a description of the individual is provided, despite the fact that the description contains no useful information.
- Representativeness bias may affect verification when we overestimate the degree to which good test results prove that our system is correct

Probability and Representativeness



- Our system is the product of either process A, which produces reliable systems, or B, which produces flawed systems
- Based on a set of test results R, what is the likelihood ($P(A|R)$) that process A built the system?
- Assume that aerospace production systems are very reliable
 - Case 1: 50% chance that we used process A [$P(A)/P(B) = 1$], and $P(R|B) = 99\%$.
 - $P(A|R) = 50.25\%$ --no significant improvement
 - Case 2: $P(R|B) = 99\%$ (very reliable process), but $P(B)$ large (very likely that a subtle flaw was introduced)
 - $P(A|R) \approx P(A)/P(B)$
 - No new information beyond our a priori estimate
- In either case, little new information was gained through the verification process, but based on the similarity between a successful system and our successful results, the tendency is to assume that we have proven that the system is reliable

$$P(A|R) = \frac{x\alpha}{1+x\alpha}, \text{ where}$$

$$x = \left(\frac{P(R|A)}{P(R|B)} \right), \text{ and}$$

$$\alpha = \left(\frac{P(A)}{P(B)} \right)$$

Positive Test Strategies

- A positive test strategy (PTS) focuses on tests that would confirm the going-in hypothesis
 - Efficient, since the number of cases is limited
 - An indispensable part of system verification
 - However, a PTS can contribute to a confirmation bias and support a representativeness bias
- Classic experiment: Wason's number series game
 - Given that (2,4,6) satisfies a rule for selecting an order series of 3 numbers, how quickly can you determine the rule?
 - Results show that participants who employ a PTS take longer to get the right answer, while those who attempt to falsify their hypothesis converge more quickly
- Implications for system verification are clear: attempts to prove the system is broken will yield more useful results
- However, it can be very challenging to identify those tests

Availability

- Availability bias occurs when estimates of the likelihood of an outcome are based on how easily instances of the outcome can be recalled
 - “We’ve had problems with that on almost every project I worked on” (possible overestimate of likelihood)
 - “In the last 30 years we’ve never had a problem with that.” (Possible underestimation)
- A significant problem in system verification
 - Obvious failure scenarios are routinely tested
 - Almost by definition, every major system failure is the result of events occurring in the deployed system that were not foreseen in development or verification (or were judged too unlikely to be worth designing/testing for)
- Complex systems have complex internal and external interactions
 - The potential space is almost infinite, but verification resources are limited
 - Cost, schedule, and social pressure can exacerbate the problem
- Mitigations can include
 - Using Monte Carlo test methods
 - Red Teams/Brainstorming sessions
 - Fostering a culture that takes “far out” failure scenarios seriously

Overconfidence

- Often (and rightly so) presumed
- Affects all domains
- A classic study highlights the problem:
 - A group of trained experts are provided with progressively more information about a situation
 - After each increment they record their conclusions
 - Findings:
 - Final accuracy (28%) only slightly better than chance (20%)
 - Accuracy did not improve significantly after each successive round
 - However, confidence in their conclusions increased after each round
- The implications for system verification are disconcerting
 - Challenges the notion that more testing is better
 - Implies that decisions to deploy may be based on “feeling good” rather than provable assertions

Anchoring and Adjustment

- Introduced in 1974 in the ground breaking paper by Kahneman & Tversky
 - “adjustments are typically insufficient”
- An initial value can exert an undue influence on the result
 - Classic example: “How long is the Amazon River?”
 - Starting estimates and other numerical input systematically lead test participants to incorrect answers
 - Most studies focus on quantitative anchoring, but the effect is probably more general
- Engineers and scientists are taught to do this (perturbation theory)
 - Becomes habit, may be used inappropriately
- Relevance to verification
 - An initial guess as to how likely a failure scenario is could lead an entire team to overlook an important case, or to expend resources unnecessarily
 - May also affect ability to conceive of new test scenarios or test approaches
- Avoiding the casual use of guesstimates can mitigate the effect

Debiasing

- Despite significant study, few successful strategies to combat bias at the individual level have been found
- Self-debiasing not reliable, and can even exacerbate the problem
- Difficult or impossible to “calibrate out” a bias after the fact
- Awareness, motivation, and self-control improve the odds of success
- Awareness has a positive effect
 - Detail on the direction and magnitude can help
 - Personalized information on a person’s measured bias is even more helpful
- The “contamination” approach
 - Set of questions can be used to determine whether a decision process has been contaminated by bias
 - When due to biasing information, avoiding exposure can be effective
 - Once contamination has occurred, it is much harder to counteract than most people believe
 - Replacement with another person, process, or tool may be the best approach

Conclusion

- Bias can effect system verification in many ways
 - Overconfidence → Questionable decisions to deploy
 - Availability → Inability to conceive critical tests
 - Representativeness → Overinterpretation of results
 - Positive Test Strategies → Confirmation bias
- Debiasing at individual level very difficult
- The potential effect of bias on the verification process can be managed, but not eliminated
 - Worth considering at key points in the process