ACQUISITION INNOVATION
RESEARCH CENTER

BAYESIAN METHODS FOR INTEGRATED TESTING

An Introduction

Justin Krometis
Virginia Tech National Security Institute

VIRGINIA TECH

Outline

- Bayesian Basics
- Computation
- Applications to Integrated T&E

The Big Idea

- Treat unknowns as statistical quantities

- Incorporate all information, including what is known before running an experiment, into the modeling via rigorous assumptions

- Update those beliefs once data is collected

Bayesian Basics

Bayes' Theorem

Relates the probability of a parameter value θ given data Y ($P(\theta|Y)$) to the probability of Y given θ and the probability of θ:

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)}$$

Note that this is a simple consequence of

$$P(\theta \cap Y) = P(\theta|Y)P(Y)$$
$$P(\theta \cap Y) = P(Y|\theta)P(\theta)$$

Bayes' Theorem (continued)

Usually $P(Y)$ is hard to calculate (requires considering all values of θ) and we really care about relative probability in θ, so we write simply:

$$P(\theta|Y) \propto P(Y|\theta)\, P(\theta)$$

Bayes' Theorem (continued)

Usually $P(Y)$ is hard to calculate (requires considering all values of θ) and we really care about relative probability in θ, so we write simply:

$$P(\theta|Y) \propto P(Y|\theta)\boxed{P(\theta)}$$

Prior

Bayes' Theorem (continued)

Usually $P(Y)$ is hard to calculate (requires considering all values of θ) and we really care about relative probability in θ, so we write simply:

$$P(\theta|Y) \propto \underbrace{\boxed{P(Y|\theta)}}_{\text{Likelihood}} \underbrace{\boxed{P(\theta)}}_{\text{Prior}}$$

Bayes' Theorem (continued)

Usually $P(Y)$ is hard to calculate (requires considering all values of θ) and we really care about relative probability in θ, so we write simply:

$$\underbrace{\boxed{P(\theta|Y)}}_{\text{Posterior}} \propto \underbrace{\boxed{P(Y|\theta)}}_{\text{Likelihood}}\underbrace{\boxed{P(\theta)}}_{\text{Prior}}$$

Bayesian Ingredients: Prior

- Probability distribution describing information known about parameters prior to gathering data
- Might be derived from:
  - Previous data, e.g.,
    - Previous phases of test
    - Digital representations of systems
    - Similar systems
  - Theory (e.g., physics-based models)
  - Subject matter expertise

Bayesian Ingredients: Likelihood

- Describes the probability of a specific test result given a set of model parameter values
  - Higher when parameter θ "matches" data Y and lower when it doesn't
- Sometimes thought of as comprising a forward model mapping the parameter space to the data space and some observational noise in measuring the data, e.g.,

$$Y = G(\theta) + \epsilon, \qquad \epsilon \sim N(0, \sigma^2)$$

yields

$$P(Y|\theta) \propto \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{1}{2\sigma^2}\left(Y - G(\theta)\right)^2]$$

Bayesian Ingredients: Posterior

ACQUISITION INNOVATION RESEARCH CENTER
VIRGINIA TECH

- Updated understanding of parameter after data are incorporated
- Describes what we can and cannot discern about model parameters from the combination of prior knowledge and new test results
  - Higher where parameter matches both data and prior understanding
  - Lower where parameter has significant mismatch with one or the other

$$\underbrace{\boxed{P(\theta|Y)}}_{\text{Posterior}} \propto \underbrace{\boxed{P(Y|\theta)}}_{\text{Likelihood}} \underbrace{\boxed{P(\theta)}}_{\text{Prior}}$$

- The posterior is the "answer" in the Bayesian setting

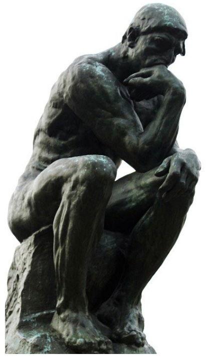
A Note about Priors

- What is the effect of the prior? Will it bias the results?
  - Yes – amount depends on how strong the prior is and how much/certain the data is
  - Large data limit: Priors are typically overwhelmed except in rare cases
- Assumptions are clearly documented
- Analysis of sensitivity to priors is common
- Frequentist Approach: Bayesian with a diffuse (e.g., uniform) prior
  - Is this really all that we know?

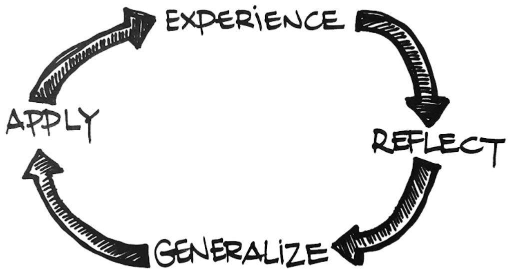
Modeling Human Understanding

Bayesian inference models how humans learn and understand

1. We start with an understanding of the world (prior)
2. We have an experience (data and likelihood)
3. We update our understanding (posterior)

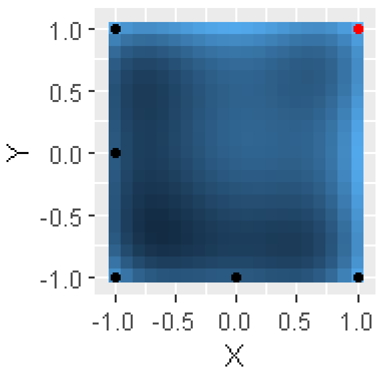


- If our understanding is vague, data can shift it significantly
- If our understanding is strong, data has limited effect



Briefly: Bayesian Design of Experiments

- Bayesian methods can be used to design tests
  - Given knowledge of system (estimates and uncertainty), where should we test next to learn as much as possible?
- Maximize expected utility: $U(d) = \int_{R\times\Theta} u(d,y,\theta)p(y|\theta,d)p(\theta)dyd\theta$
  - e.g., Bayesian D-optimality, Mutual information (Kullback-Liebler divergence)
- Similar to conventional DOE when:
  - Priors are weak
  - Data is limited
- Can be extremely computationally intensive
  - Mostly limited to low-dimensional designs (e.g., 4 parameters)

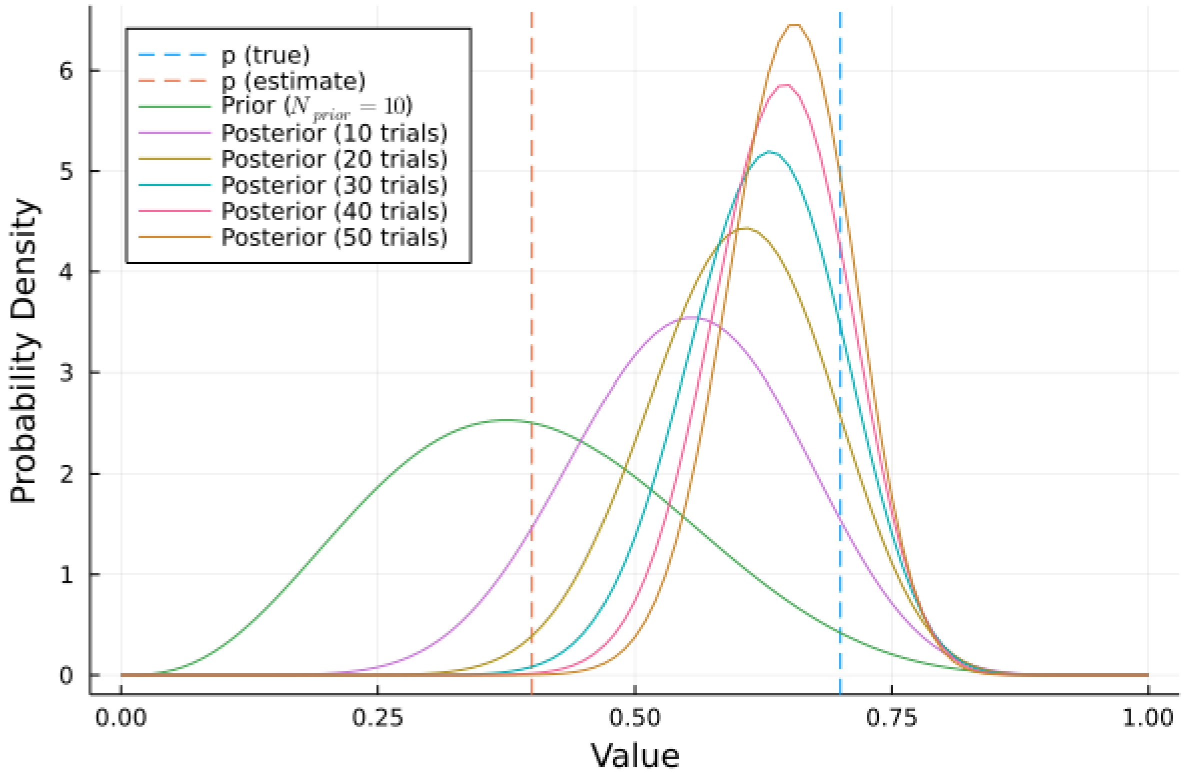


Computing the Posterior

How do we get the posterior?

- We have the equation – can we just write it down?
  - Typically not in any nice, usable form, as probability of a given θ requires knowing the normalization $P(Y)$, which we can't compute
- Exception: Conjugate priors
  - Yield posterior distribution in same family as prior distribution
- Example:
  - Estimate the probability of success θ of a Bernoulli trial
  - Prior: $Beta(\alpha, \beta)$ for some choice of α and β
  - Likelihood: Binomial distribution
  - Observing s successes and f failures, yields the posterior: $Beta(\alpha + s, \beta + f)$

Beta-Binomial Conjugate Prior Example



Markov Chain Monte Carlo (MCMC)

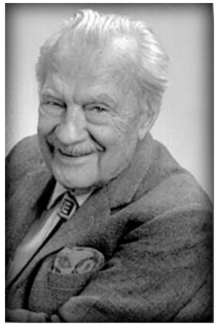
- Idea:
  - Find a Markov Chain with the target distribution as the invariant (steady-state) distribution
  - Under certain conditions, if we draw enough samples from the chain, we get samples from the (approximate) target distribution
  - Use samples to compute quantities of interest, e.g., mean, standard deviation, quantiles, credible intervals, etc
- Does not require knowing the normalization constant

The Original MCMC: Random Walk

To sample from a target density $\pi(\theta)$:

- Start with a sample θ_0
- Sample from a normal $\tilde{\theta} \sim N(\theta_0, \sigma^2)$
- Evaluate $\pi(\theta_0)$ and $\pi(\tilde{\theta})$
- Rule: Always go uphill, sometimes go downhill
  - If $\pi(\tilde{\theta}) \geq \pi(\theta_0)$, "accept" $\theta_1 = \tilde{\theta}$
  - If $\pi(\tilde{\theta}) < \pi(\theta_0)$
    - "Accept" $\theta_1 = \tilde{\theta}$ with probability $\pi(\tilde{\theta})/\pi(\theta_0)$
    - Otherwise "reject" and stay in the same place: $\theta_1 = \theta_0$
- Repeat starting from θ_1

MCMC: Continued

- Many methods, e.g.:
  - Gibbs samplers
  - Metropolis-Hastings samplers
- Implementation can be complicated depending on the method
- Nice visualizations of various methods can be found here:

  https://chi-feng.github.io/mcmc-demo/
- Fortunately, some standard implementations exist, e.g.,
  - JAGS ("Just Another Gibbs Sampler"), available via R package `rjags`
  - Stan, available via R package `rstan`

Terminology: Confidence vs. Credible Intervals

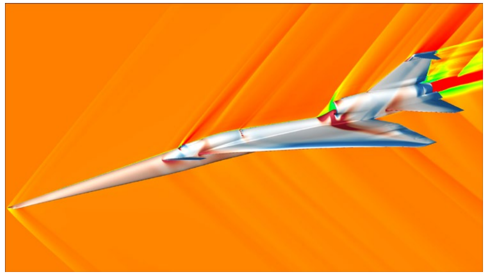
- Frequentist statistics: Confidence interval is an interval $[a, b]$ associated with results of an experiments such that the true value of the parameter θ will be in $[a, b]$ in $\gamma = 1 - \alpha$ (e.g., 95%) of such experiments
  - Example statement: "In 95% of experiments, $\theta \in [a, b]$"
- Bayesian equivalent: Credible interval is the interval $[a, b]$ such that the posterior probability $\mathrm{P}(\theta \in [a, b] | Y) = \gamma = 1 - \alpha$
  - Example statement: "With 95% probability, $\theta \in [a, b]$"
  - MCMC: Credible intervals can be estimated from quantiles of the samples

Bayes for Integrated T&E

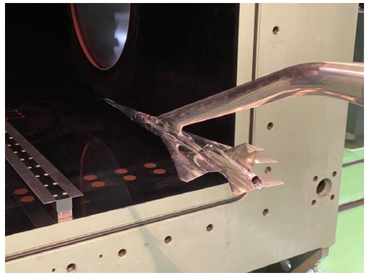
Data Sources

- There is a lot of data generated as a technology is developed

Historical Data
SME Knowledge



Model Data



DT Data

OT Data

Operational
Performance

- Can we use Bayesian inference to build an integrated picture of behavior?

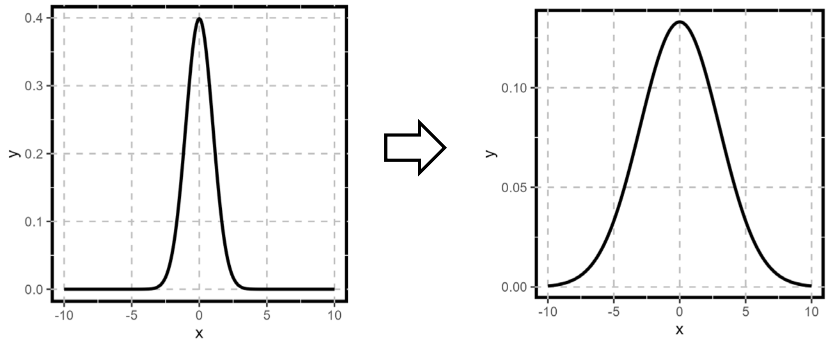
A Nice Feature

- Consider two datasets, Y_1 and Y_2, that were collected in the same fashion
- Under many (most?) likelihood models…
  - Computing the posterior using both Y_1 and Y_2 as a single dataset
  - Computing the posterior associated with Y_1, using that as the prior, and computing the posterior associated with Y_2

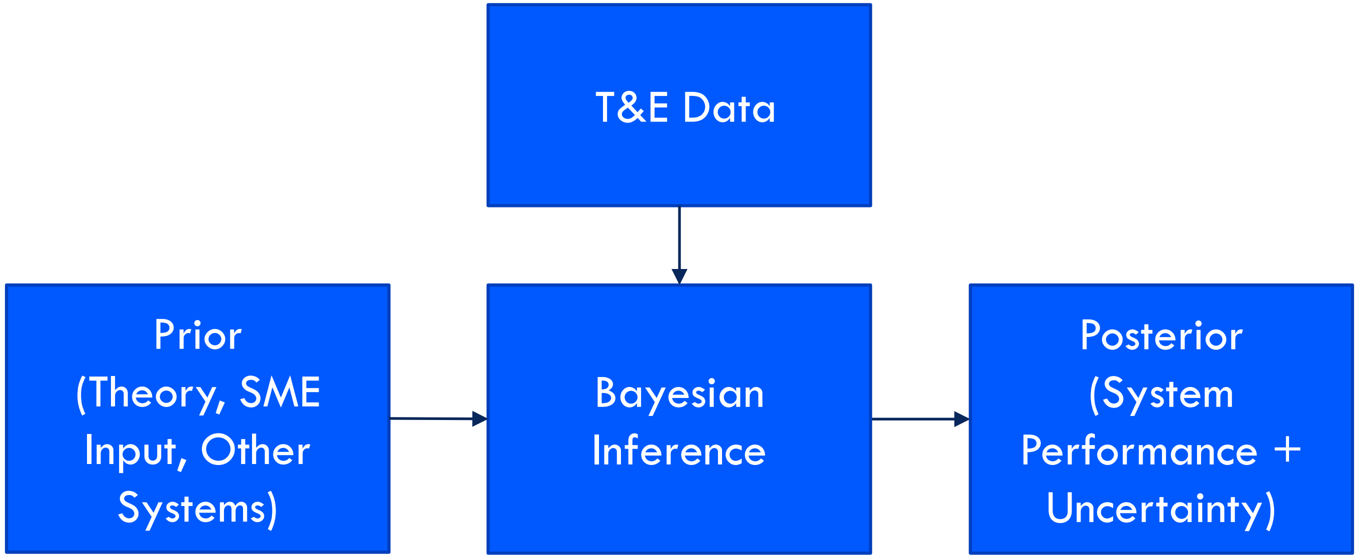
  …produce the same answer
- We can iterate inference to build understanding as data accumulates

Changes in Data/Context

- What if something changes in between data collections?
  - System updates
  - Environment or context changes
- Then posterior from one dataset can be modified before using as prior for subsequent analysis
  - Example: “Downweighting” a posterior to increase its uncertainty – trust the previous data less than the new data



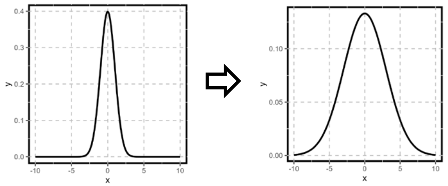
Single Test Phase



Single Test Phase

T&E Data

Prior (Theory, SME Input, Other Systems) → Bayesian Inference → Posterior (System Performance + Uncertainty)



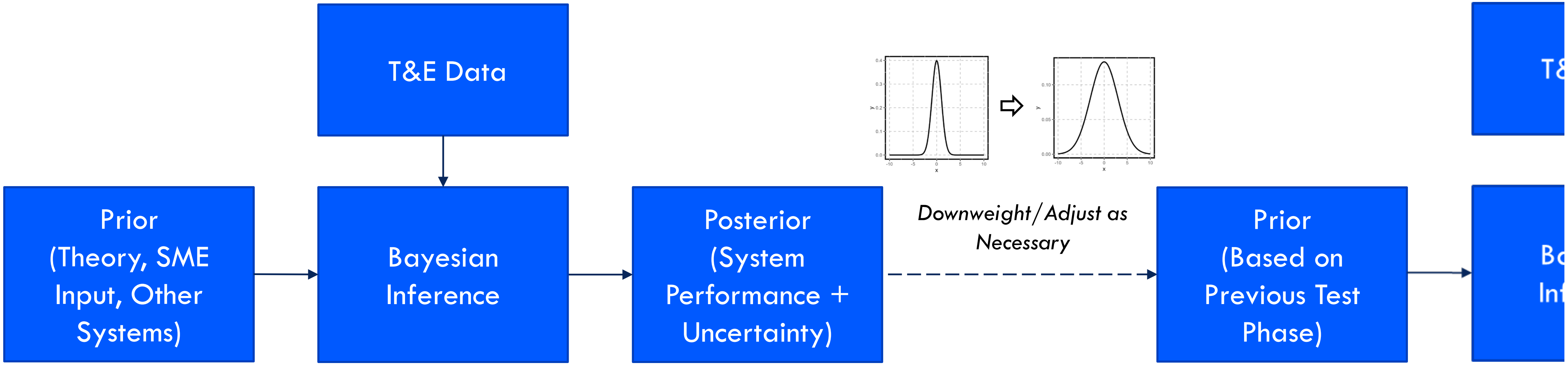
Downweight/Adjust as Necessary

Prior (Based on Previous Test Phase)

Iterated Inference

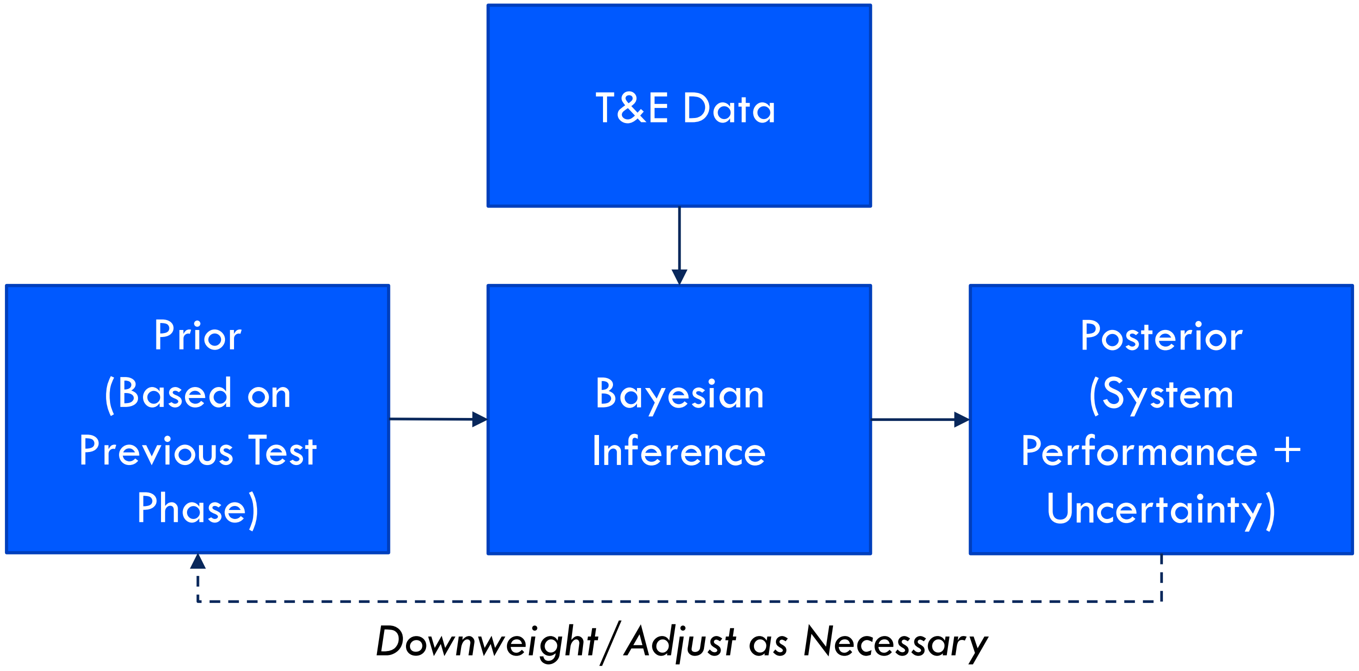
Single Test Phase

Next Test Phase

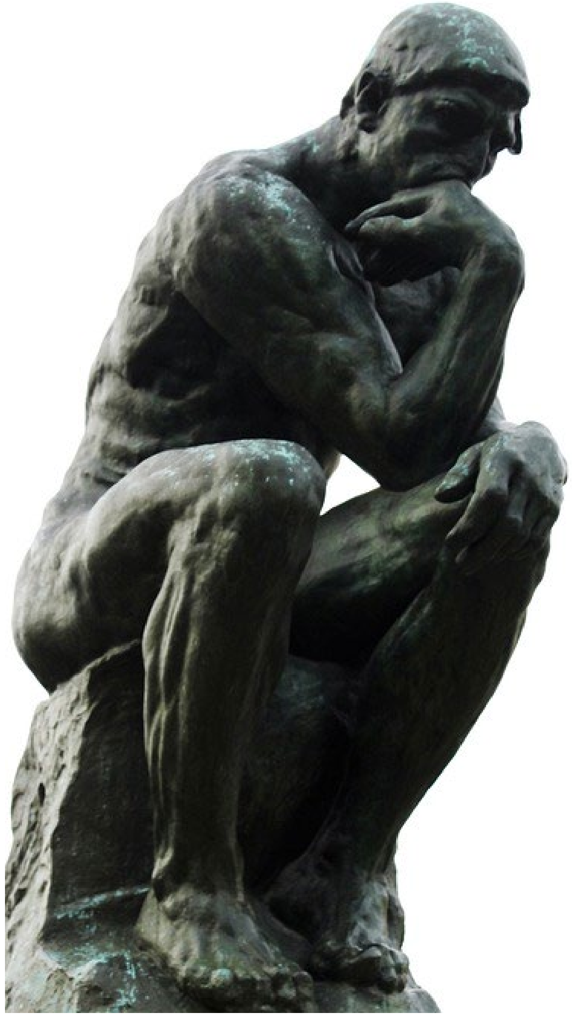


Not Cleared for Public Release

Multiple, Integrated Test Phases



Questions?



Not Cleared for Public Release