

A photograph of two tropical cocktails, likely Piña Coladas, served in large, round glasses with long stems. Each cocktail is garnished with a slice of pineapple and a small sprig of mint. They are placed on a sandy beach. In the background, the ocean is visible under a sky filled with warm, orange, and yellow hues from a setting sun.

Bayes on the Beach 2024

# COOLANGATTA GOLD

Bayesian mixture modelling...  
of delicious mixtures for Bayesians.



ARC Training Centre in  
Data Analytics for Resources & Environments



**Australian Government**  
**Australian Research Council**

## Acknowledgement of Country

Today we begin by acknowledging the Traditional Custodians of the lands and waters of the Gold Coast; the Yugambeh people, and pay our respects to their Elders past, present and emerging.

As we share our own knowledge, teaching, learning and research practices within this conference, may we also pay respect to the knowledge embedded forever in the Aboriginal Custodianship of Country.



THE UNIVERSITY OF  
SYDNEY



UNSW  
SYDNEY



THE UNIVERSITY OF  
WESTERN  
AUSTRALIA



MACQUARIE  
University  
SYDNEY-AUSTRALIA



Australian  
National  
University



**DARE**

ARC Training Centre in  
Data Analytics for Resources & Environments



**Australian Government**  
**Australian Research Council**



# Bayes on the Beach Challenge: Coolangatta Gold

Dr. Travis Stenborg  
Research Engineer  
DARE / University of Sydney



THE UNIVERSITY OF  
SYDNEY



**UNSW**  
SYDNEY



THE UNIVERSITY OF  
WESTERN  
AUSTRALIA



**MACQUARIE**  
University  
SYDNEY-AUSTRALIA



Australian  
National  
University



# Coolangatta Gold

## Ironman Race

- Iconic Gold Coast surf lifesaving race.
- Organised by Surf Lifesaving Australia.

## Australian Film

- 1984 film starring actors and real ironmen.

## Bayes on the Beach Cocktail

- A fictional cocktail invented for this challenge!
- Based on a particular specialty Martini.

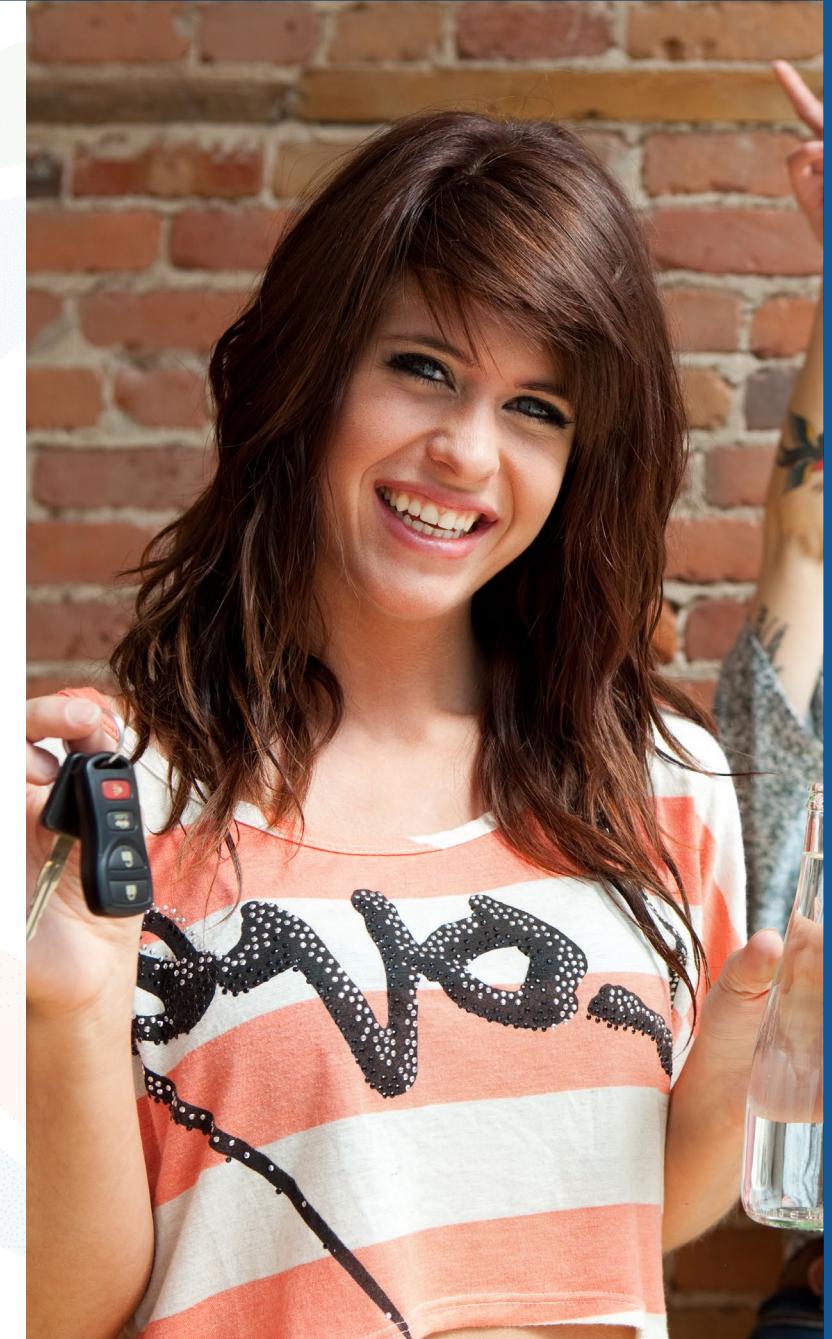




# Alcohol Awareness

## Moderation

- Alcohol can be fun, but...
- Consider your health.
- Drink responsibly.





# Prerequisites

## R + Stan

- Ensure R is installed.

<https://cran.csiro.au/>

- Invoke Stan via the rstan package.

<https://mc-stan.org/users/interfaces/rstan>

- RStudio may be helpful too.

<https://posit.co/download/rstudio-desktop/>





# Secret Ingredient

## Coolangatta Gold

- 1.5 ounces *SECRET INGREDIENT*.
- 0.5 ounces yellow Chartreuse.
- 1 dash orange bitters.
- Garnish: pineapple slice.

## Challenge

- Find the secret ingredient.
- Use Bayesian mixture modelling.





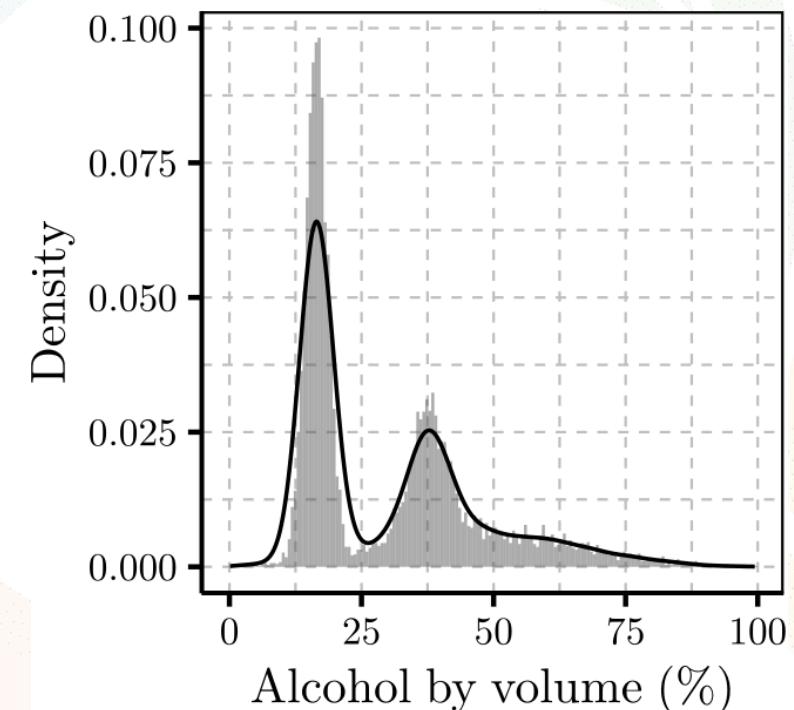
# Gaussian Mixture

## Cocktail Testing

- A mixture of 10,000 test cocktails.
- They used a common base ingredient set.
- Each one also had a secret ingredient added.
- The secret (alcoholic) ingredient could vary between test cocktails.
- Small number of distinct secret ingredients.

## Imprecise Alcohol by Volume

- Different **interingredient** means.
- Dispersed **intraingredient** values.



Probability density function of the alcohol by volume for the secret ingredient in 10,000 test cocktails.

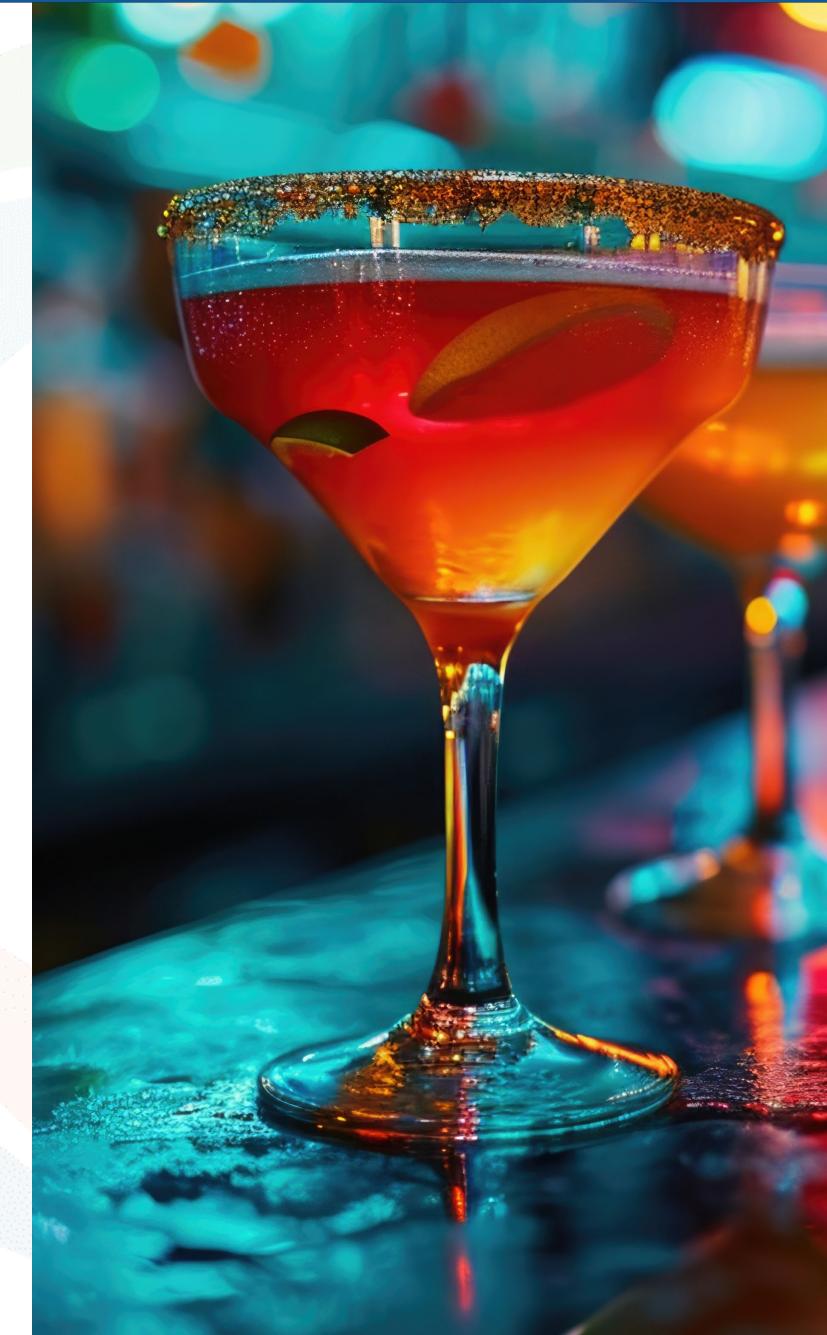


# Possible Secret Ingredients

Ingredients *potentially* in the mixture model.

Ingredient	Relative $\sigma$ (alcohol by vol %)
Gin	8.3
Pisco	1.0
Sherry	1.7
Vermouth	5.7
Vodka	2.3

The secret ingredient selected from amongst the candidates was the one with greatest mean alcohol by volume. What's the secret ingredient?





# Problem Solving Process

## Assumptions

- Gaussian mixture model.
- $\leq 5$  components.

## Please Answer...

- How many components are there?  
(And how did you determine the number?)
- What are their means, standard deviations and mixture weights? (And uncertainties?)
- What's the secret ingredient?





# Solution Rigour

## MCMC Convergence

- Describe your convergence assessment method(s).
- Traditional  $\hat{R}$ ?

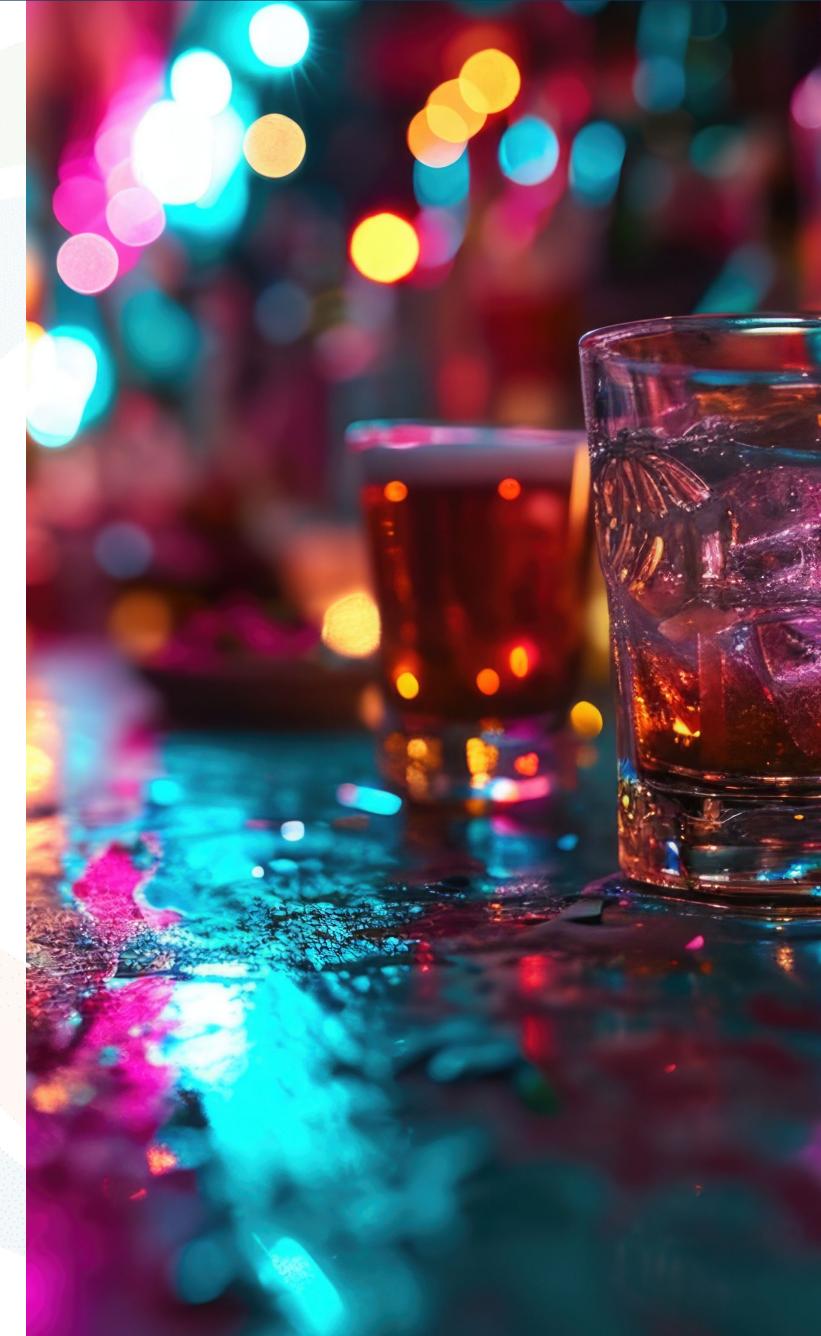
Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. *Stat Sci.* 1992;7(4):457-72.

- Split- $\hat{R}$ ?

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. 3rd ed. Chapman & Hall/CRC; 2013.

- Is there something even better?

Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner PC. Rank-Normalization, Folding, and Localization: An Improved  $\hat{R}$  for Assessing Convergence of MCMC (with Discussion). *Bayesian Anal.* 2021;16(2):667-718.





# Solution Rigour

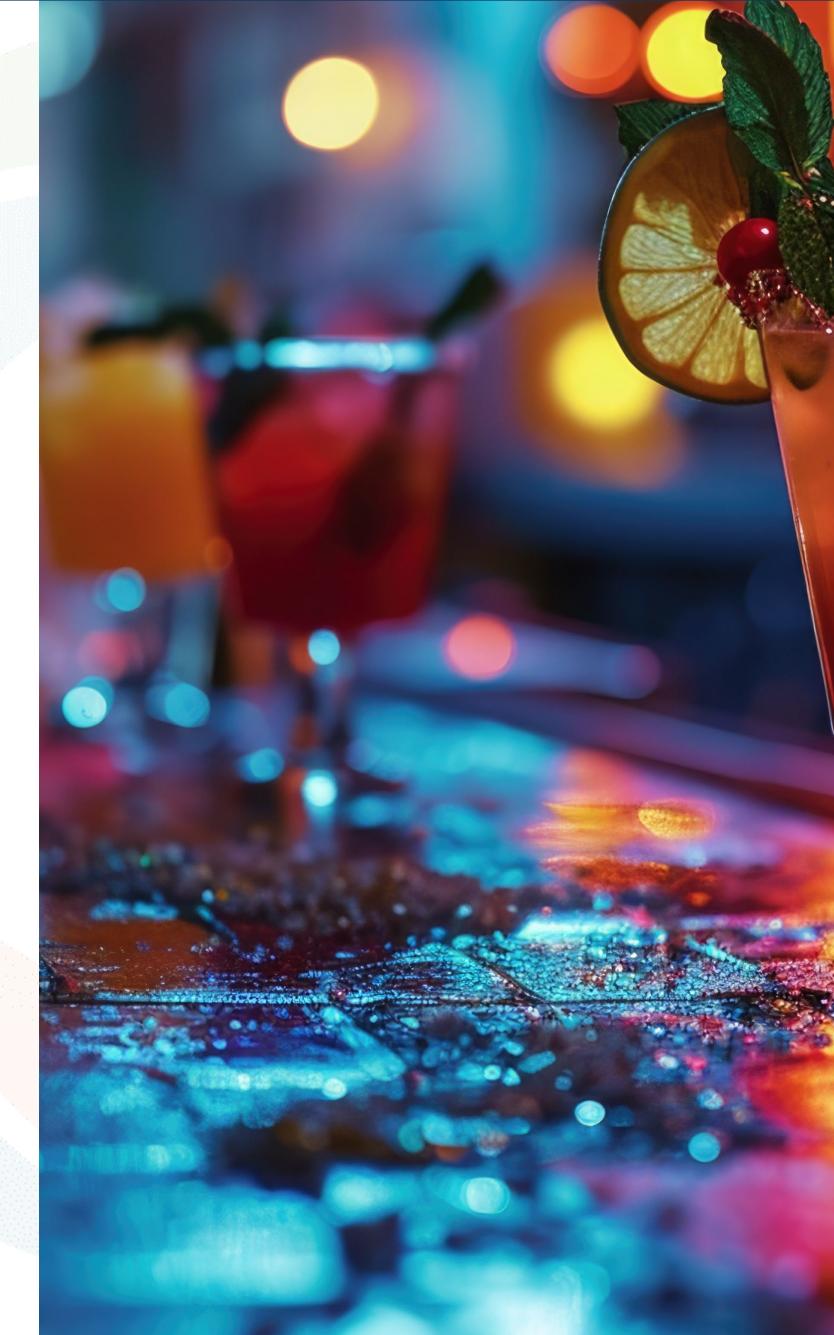
## MCMC Convergence

- What about effective sample size?  
(Is Markov chain convergence uniform across parameter space?)

Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner PC.

Rank-Normalization, Folding, and Localization: An Improved  $\hat{R}$  for Assessing Convergence of MCMC (with Discussion). *Bayesian Anal.* 2021;16(2):667-718.

- Have you considered a graphical metric?





# Performance Tuning

## Solution Speed

- Points will be awarded for fast code.
- Efficient choice of cores / chains to use is typically a function of model complexity and data volume.
- Can that choice be automated?

Stenborg TN. Adaptive MCMC parallelisation in Stan. In: MODSIM2023, 25th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand; 2023. p. 913.  
[Extended Abstract].

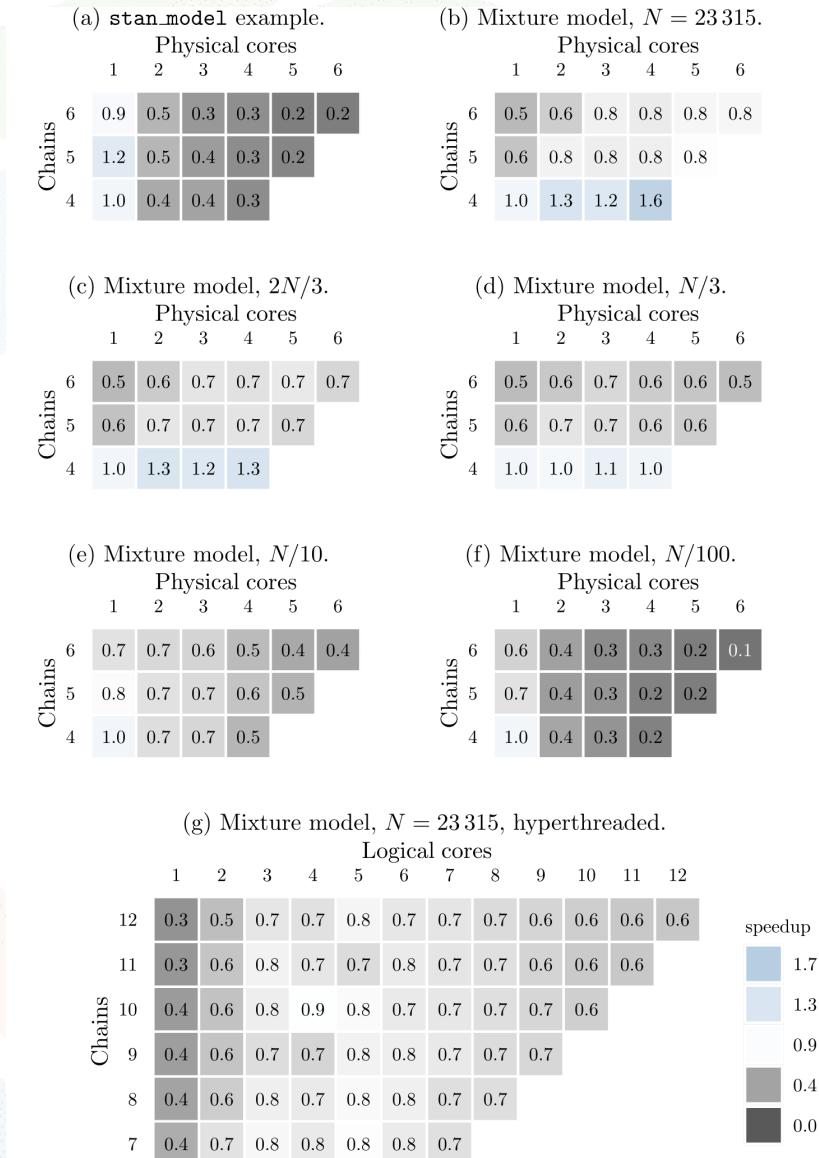


Figure 14: **MCMC parallelisation** and Stan model inference speedup. Sampling used 50 warmup and  $50/m$  post-warmup iterations/chain.



**DARE**

ARC Training Centre in  
Data Analytics for Resources & Environments



**Australian Government**  
**Australian Research Council**



<https://github.com/dare-centre/BOTB-2024-data-challenge>



THE UNIVERSITY OF  
SYDNEY



**UNSW**  
SYDNEY



THE UNIVERSITY OF  
WESTERN  
AUSTRALIA



MACQUARIE  
University  
SYDNEY-AUSTRALIA



Australian  
National  
University