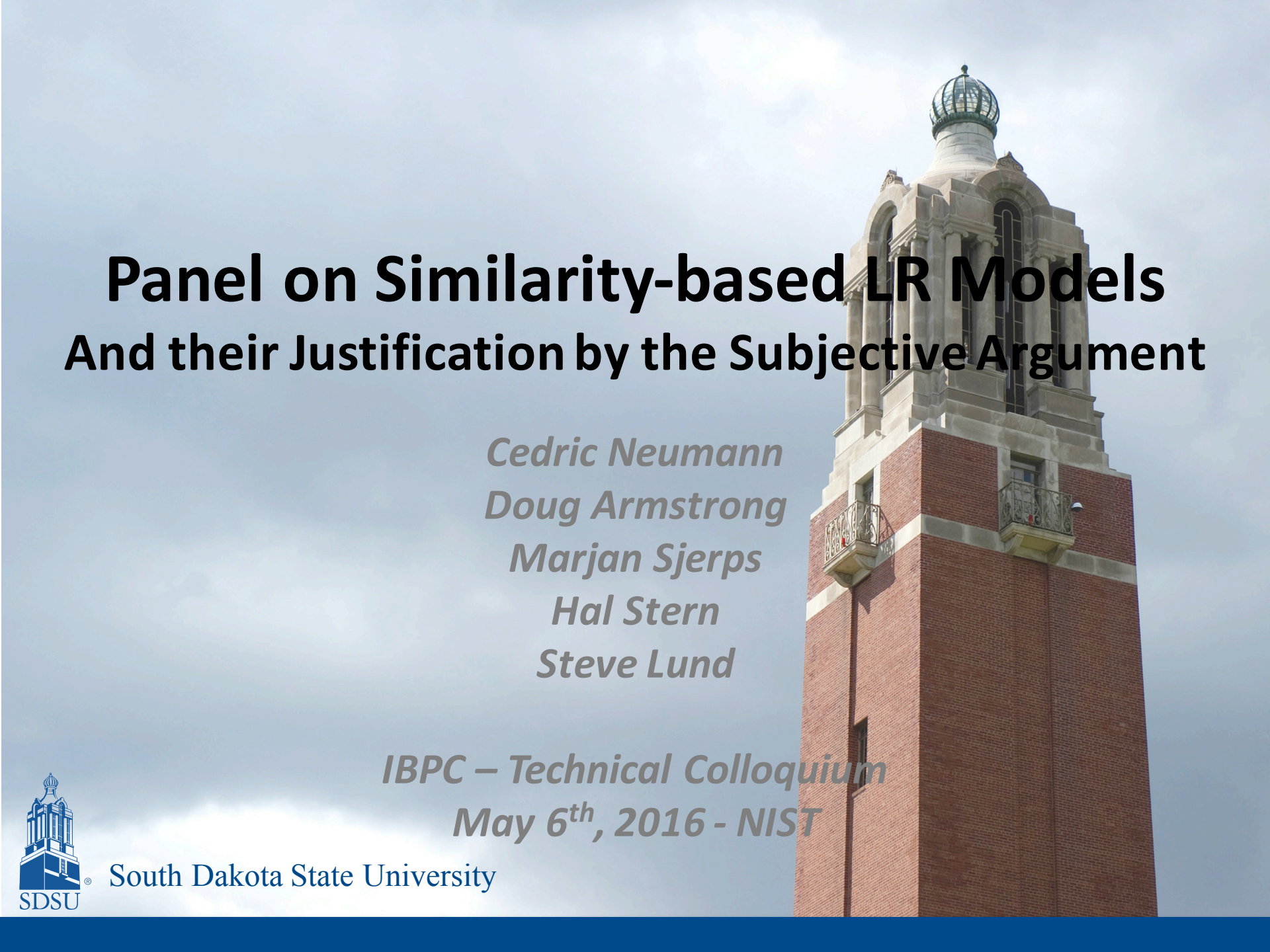

Panel on Similarity-based LR Models

And their Justification by the Subjective Argument

Cedric Neumann

Doug Armstrong

Marjan Sjerps

Hal Stern

Steve Lund

IBPC – Technical Colloquium

May 6th, 2016 - NIST

SDSU

South Dakota State University

Introduction

- Many researchers are concerned with the development of models to quantify the weight of forensic evidence
- These models take the form of a *likelihood ratio* or a *Bayes factor*
  - Some researchers make a distinction between these two terms, some don't

Introduction

- The *likelihood ratio* or the *Bayes factor* consider the evidence under two mutually exclusive propositions:
  - One representing the arguments of the prosecution
  - One representing the arguments of the defense

Introduction

- Observations made on forensic evidence are, in most cases, multivariate
  - Very high-dimension
  - Heterogenous
  - Not independent
- Extremely difficult (who said impossible ? ☺) to model in the feature space in a rigorous manner

Introduction

- Starting in the late 1990's and early 2000's some researchers, based on research in biometry (access control and database searches) proposed to use similarity measurements as a dimension reduction technique
- And then quantify the weight of the similarity measure

Introduction

- Example:
  - A shoe impression is found at a crime scene
  - It is compared to the shoe of a suspect
  - A metric measures their similarity δ
  - The probability of observing δ given that the two impressions originate from the same source is compared to the probability of observing δ given that the two impressions do not.

Introduction

- Example (continued):
  - The probability of observing δ given that the two impressions originate from the same source is compared to the probability of observing δ given that the two impressions do not.
  - Note that at this point, our weight of evidence is the same whether the trace impression was made by a shoe with a rare sole pattern or not
    - **Only δ matters…**

Introduction

- Several “similarity-based” models have been proposed in the past decade:
  - “suspect-anchored”
  - ”trace-anchored”
  - “non-anchored”

Introduction

$$SLR = \frac{\Pr(\delta(X,Y)|H_p)}{\Pr(\delta(X,Y)|H_d)}$$

$$SLR = \frac{\Pr(\delta(X,Y)|X,H_p)}{\Pr(\delta(X,Y)|X,H_d)} \frac{\Pr(X|H_p)}{\Pr(X|H_d)}$$

$$SLR = \frac{\Pr(\delta(X,Y)|Y,H_p)}{\Pr(\delta(X,Y)|Y,H_d)} \frac{\Pr(Y|H_p)}{\Pr(Y|H_d)}$$

Introduction

- Several “similarity-based” models have been proposed in the past decade:
- They all lead to different weights for the same evidence
- Some researchers have proposed to “calibrate” these models to reduce the rates of “error”

Introduction

- Ultimately,
  - These methods appear very ad-hoc
  - The methods are justified by the Subjectivity argument that

"Since probabilities are measures of personal belief, the SLR (which ever is chosen, or however it is calibrated) is merely an expression of the belief of the researcher and therefore perfectly valid"

Introduction

- My fear is that we are replacing the 1990's

  "it's him and you need to trust me because I have been doing that job for 30 years"

- By the 2010's

  "it's my LR and you need to trust me because I have been a statistician for 30 years"

Panel

- Prof. Hal Stern
  - Dean of Donal Bren School of Information and computer Sciences, and prof. of statistics, UC Irvine.
  - PI for NIST - CSAFE
- Prof. Dr. Marjan Sjerps
  - Principal scientist, NFI
  - Prof. of forensic statistics, University of Amsterdam
- Dr. Steven Lund
  - Statistical Engineering Division at NIST
- Doug Armstrong
  - PhD student at South Dakota State University
  - NIJ Research Fellow.

Panel

- Rules
  - 3 questions
  - 10 minutes to comment on all three questions
  - 5 minutes to respond to other panelists
  - Questions/comments from public
- I will strictly enforce the timing
  - During the panel
  - Also for questions/comments from the public
    - Announce yourself
    - Be brief and to the point
      (I will interrupt you after a couple of minutes)

Panel

- 3 questions:
  - Is it acceptable to use the argument of the subjective nature of probabilities to justify *any* construction of the LR/BF?
  - Is it acceptable to "calibrate" LRs to reduce the rates of "errors"? What are the benefits/limitations of such practice?
  - Given that score-based models are controversial and feature-based models are nearly impossible to develop rigorously, what's next?