Integrating Data Types to Understand the Impact of Early-Life Experiences

Big Data Training
for Cancer Research
July 22, 2025

Hal Stern
Department of Statistics
sternh@uci.edu

UCI Department of Statistics
Donald Bren School of Information & Computer Sciences

Conte Center at UCI

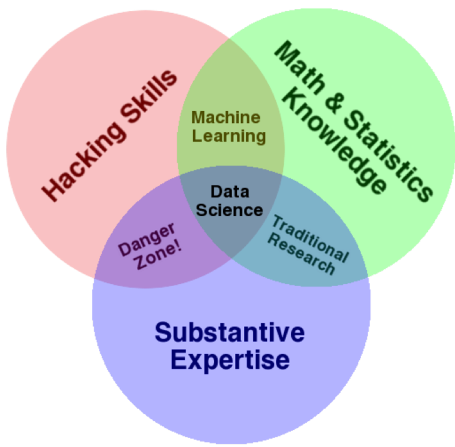
- NIMH-funded (2013-2018, 2019-2024)
  - Center studies the impact of early life experience (especially unpredictability) on adolescent/adult mental health
  - Four related projects addressing this common theme
    - Animal model (rodent) – experiments
    - Humans
      - Infants/Children (prenatal – age 7)
      - Adolescents/Young Adults (ages 16-21)
      - Marine veterans (young adults)
  - Wide range of data types
    - Emotional measures (survey questionnaires)
    - Behavior measures (e.g., risk taking tasks, videos)
    - Brain imaging
    - Genetics

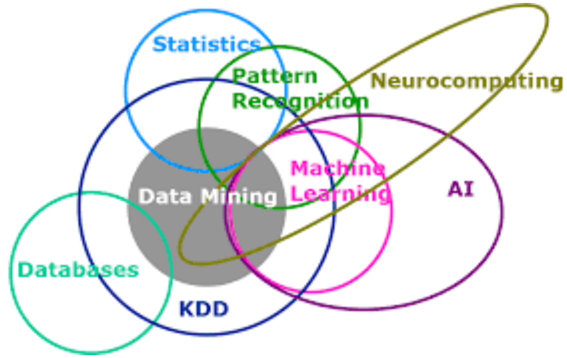
A few thoughts on "Big Data"

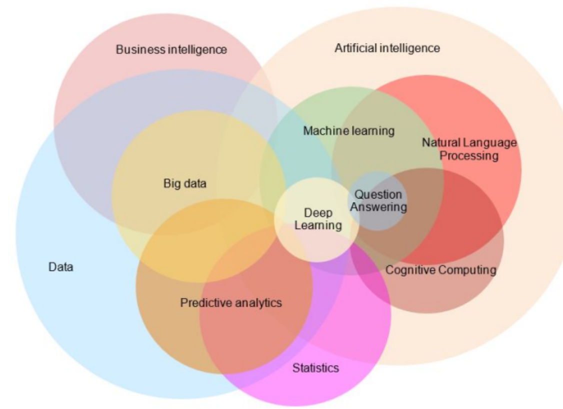
- There are many terms associated with data analysis these days. Examples include:
  - Statistics
  - Machine Learning (ML)
  - Data Science
  - Big Data
  - Artificial Intelligence (AI)
  - Deep Learning (Deep Neural Networks)
- This has proven confusing and led to many attempts to clarify …

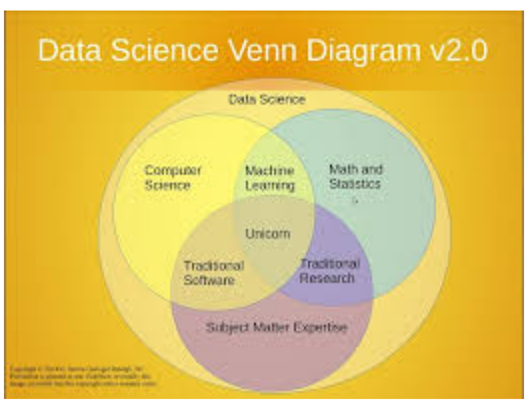
A Venn Diagram to Explain it All

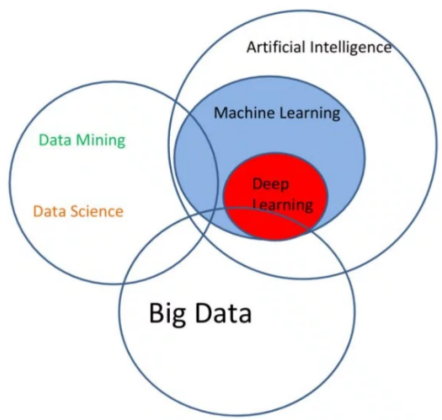
One of the first by Drew Conway

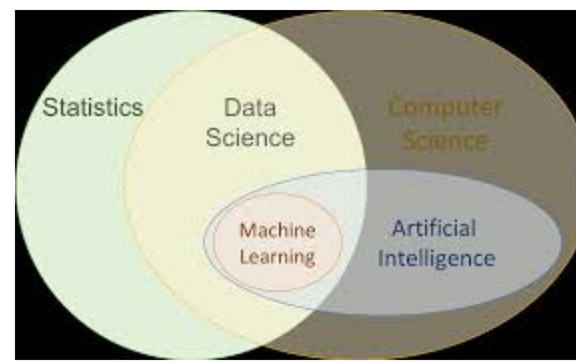






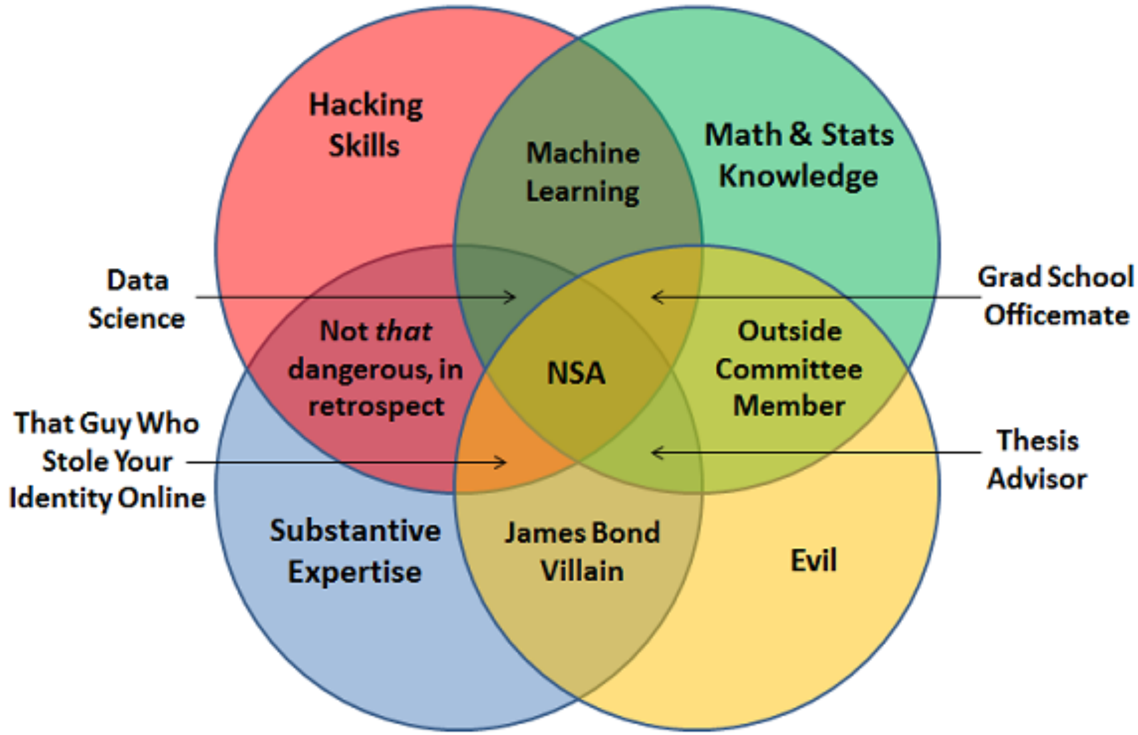






Terminology

A humorous take credited to Joel Grus:

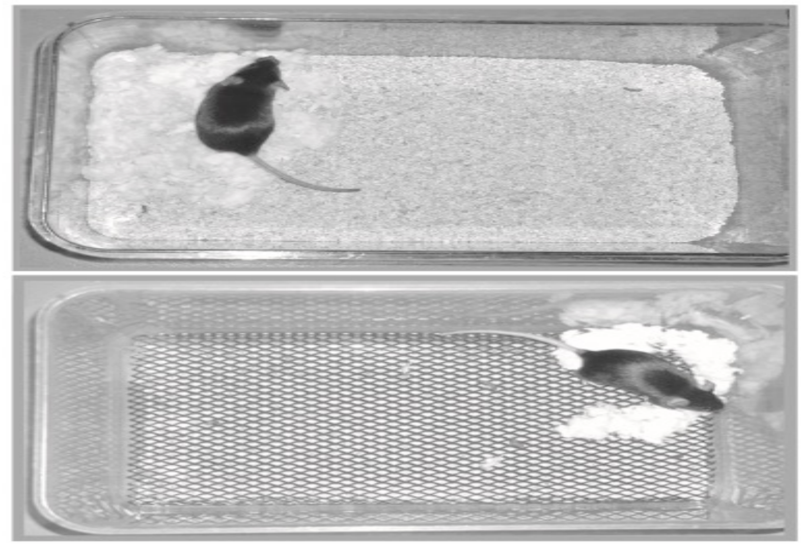


Terminology

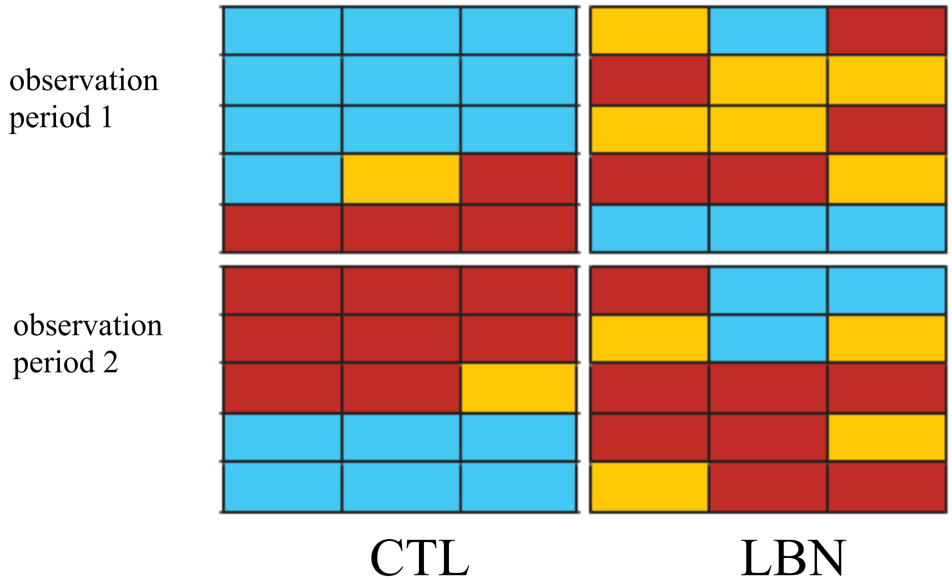
- The (stereotyped) view of AI/ML from Statistics
  - AI/ML = CS discovering the power of probability and statistical models to solve problems / analyze data
  - ML folks are not concerned about where the data come from
- The (stereotyped) view of Statistics from AI/ML
  - Statistics is focused on mathematical theories for data analysis
  - They think primarily about interpretation / testing of models
  - Can't handle very large data sets
- There are elements of truth in these stereotypes, but ….
- Main point for me is that there are a wide range of tools available to help scientists make sense of their data
- Use best tool for the task at hand

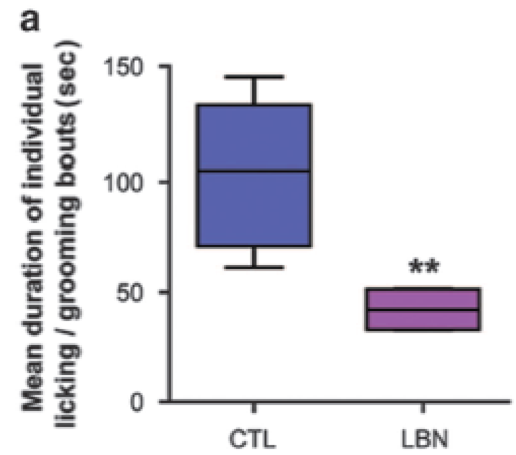
The role of unpredictability

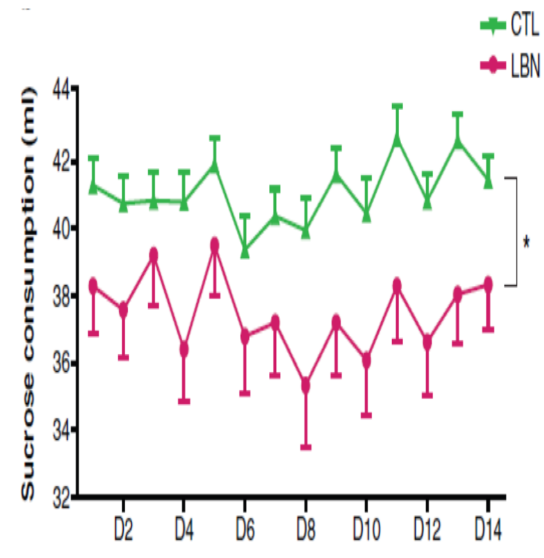
- Experimental protocol in the rodent animal model
  - Top cage=normal environment (CTL)
  - Bottom cage=limited bedding and nesting (LBN)
- Pups randomly assigned to CTL/LBN cages for postnatal days 2-9
- Then all returned to CTL environments
- Observe very different maternal behavior in the LBN cages (fragmented and unpredictable behavior)
- Offspring that spent time in LBN cages are vulnerable to emotional/cognitive problems



The role of fragmentation/ unpredictability



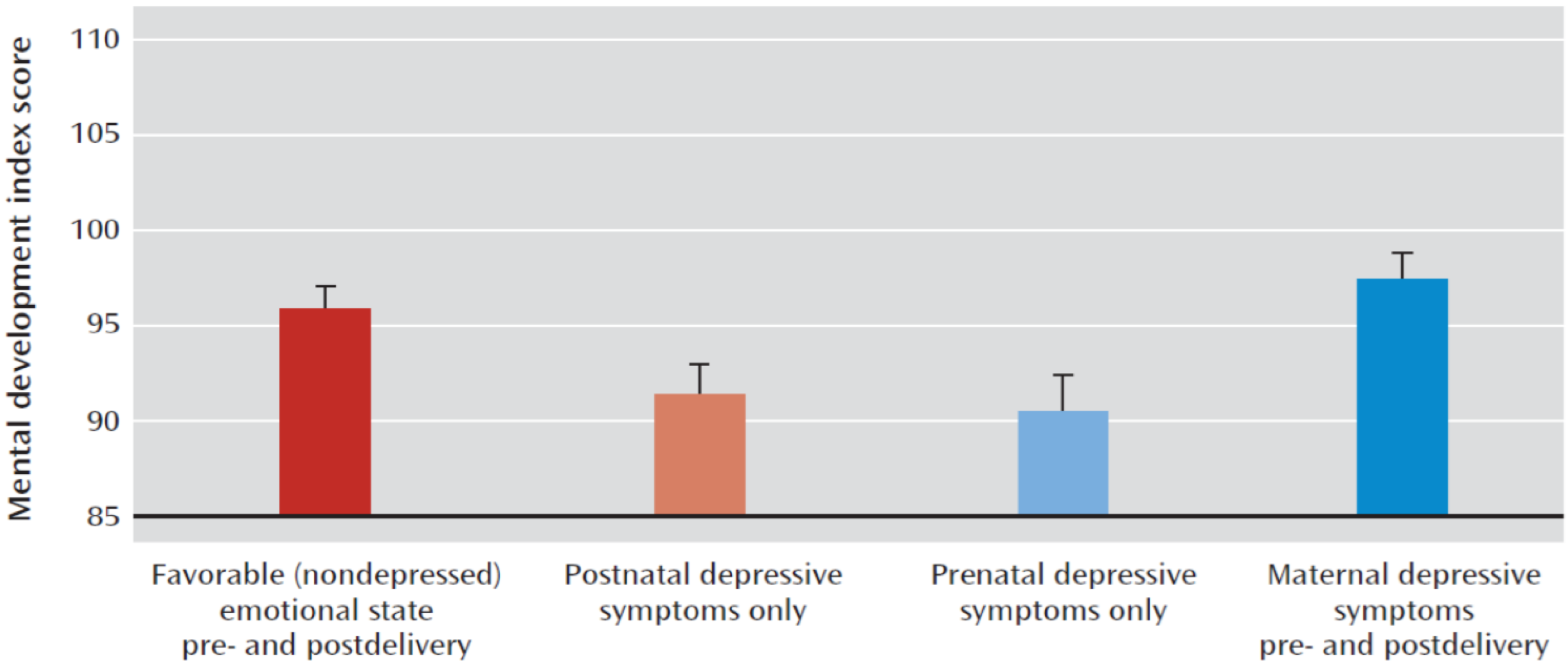




- Observed and recorded maternal behavior
- Graph shows behavior in 1minute blocks
- Different patterns of behavior, e.g.,
  - Licking/grooming occurs in shorter bouts in the LBN cages
- Different outcomes as pups mature, e.g., LBN group consumes less sucrose (a preferred item)

But does this happen in humans?

- Pre-natal and post-natal maternal questionnaires on depressive symptoms
- Median split both measures (low symptoms vs high symptoms)
- Examine child mental development (index measured at 1 year of age)
  - Children experiencing consistent maternal environment (symptoms) outperform those with inconsistent environments



UCI Department of Statistics
Donald Bren School of Information & Computer Sciences

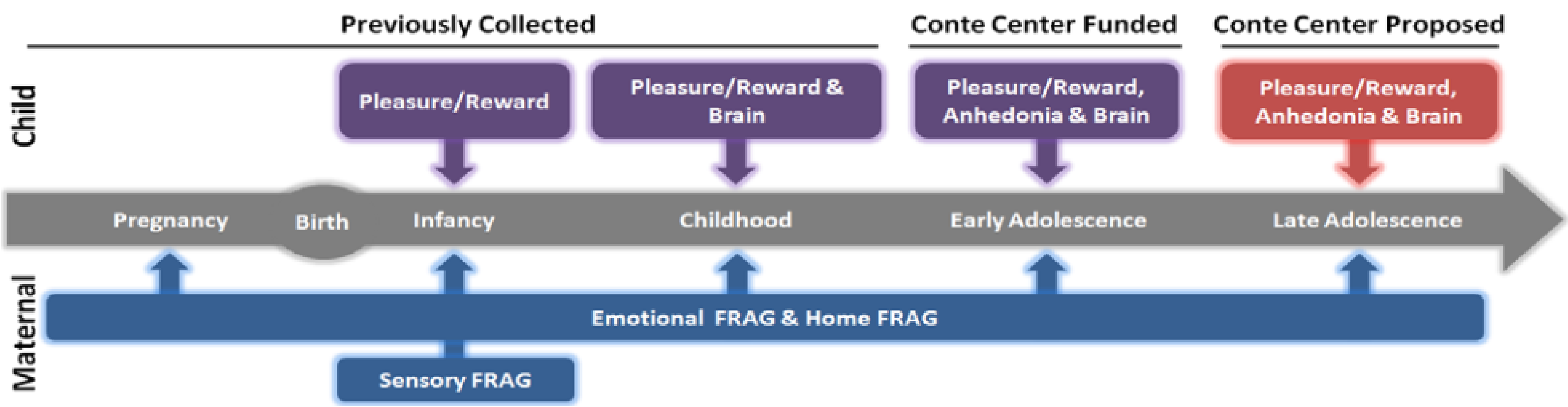
Conte Center goals

- Can we characterize fragmentation and unpredictability of early life environment using the same or similar measures across species
- Study the association between unpredictability and child, adolescent, and ultimately adult outcomes
- Use rodent models to try and understand the mechanism through which this association may develop
  - Brain imaging
  - Genetics (epigenetics)
- Try to validate mechanistic theories by examining human data

Conte Center human studies

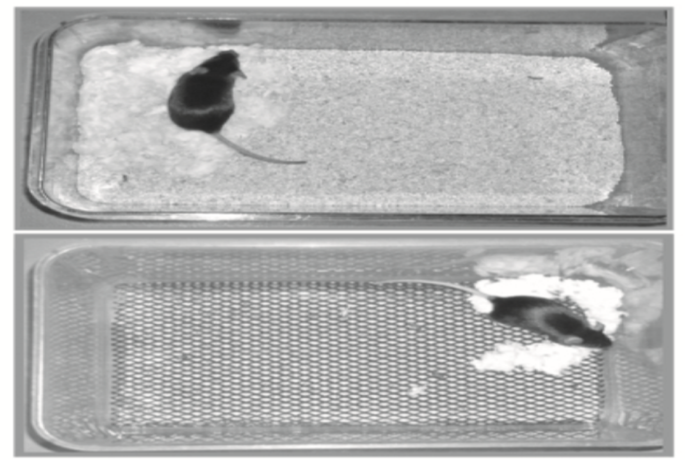
- Conte Center 1.0 leveraged an existing cohort of mother-child dyads and recruited a second cohort of mother-child dyads
- Extensive data collection
- Conte Center 2.0 is collecting additional data on both cohorts and also studying a cohort of Marines

Fig 1. **Project 3 Timeline of Assessments**



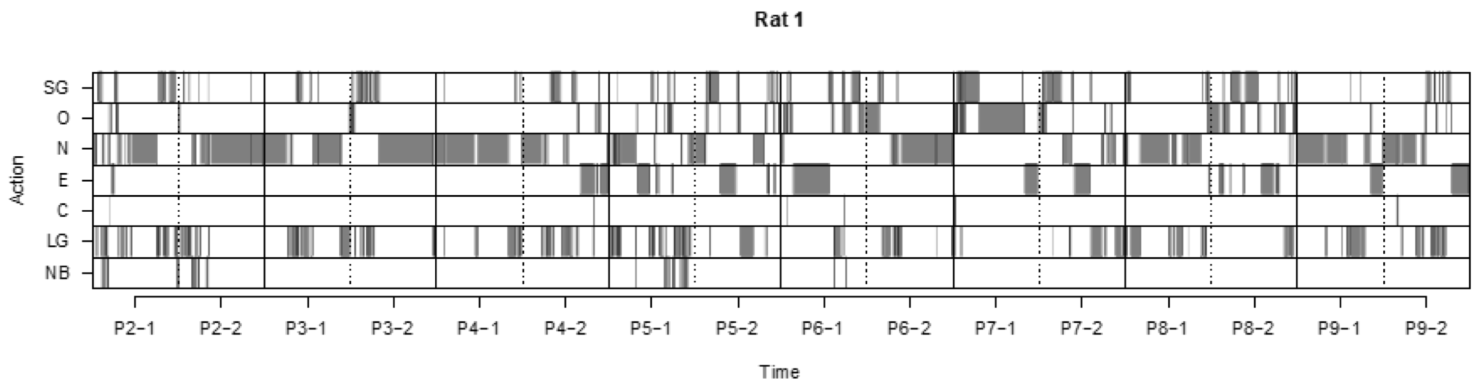
Assessing predictability of mom

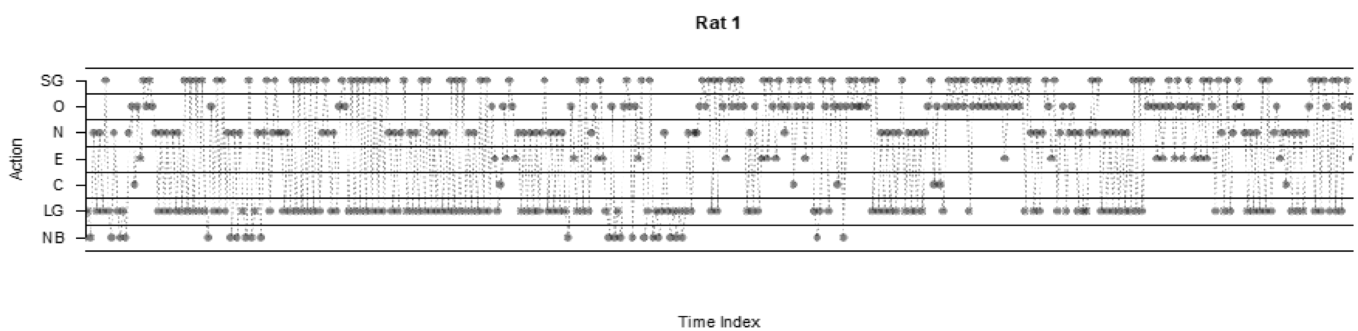
- One common data type across species is behaviors recorded during videotaped interactions between mom/offspring
  - Rodents
    - Observed for 50-minute periods (2x/day for 8 days)
    - Records of start and stop of different behaviors (licking/grooming, carrying, eating, nursing, nest building, off pups, self-grooming)
  - Humans
    - Observed 10-minute play sessions of mother and child (at 6mo, 12mo)
    - Records of many different behaviors
    - Focus on sensory input (auditory, tactile, visual)



Assessing predictability - rodents

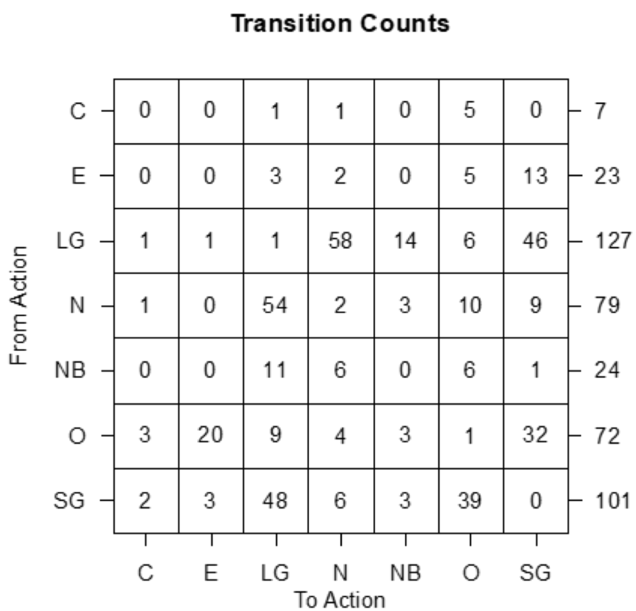
- Rodents
  - Observed for 50-minute periods (2x/day for 8 days)
  - Records of start and stop of different behaviors (licking/grooming, carrying, eating, nursing, nest building, off pups, self-grooming)

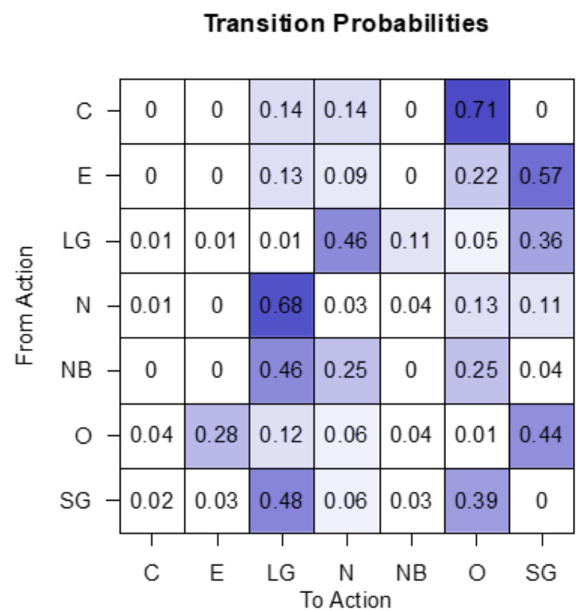




Assessing predictability - rodents

- Rodents





Entropy

- Entropy (Shannon entropy) is a measure of randomness / unpredictability
- Consider a random quantity with four possible outcomes (a, b, c, d)
- We see 10 observations of this random quantity:
  - Example 1:   b, a, a, c, b, a, b, d, c, c
- And then we see 10 observations from a second random variable with the same possible outcomes
  - Example 2:   a, a, a, a, a, a, a, a, a, a
- And then a 3rd random variable
  - Example 3:   d, c, c, d, d, d, d, c, d, c
- These three examples have very different behavior

<u>Entropy is one way to characterize the differences</u>

- Example 1 – quite random (30% a, 30% b, 30% c, 10% d) → entropy = 1.90
- Example 2 - perfectly predictable (100% a) → entropy = 0.00
- Example 3 - somewhat predictable (0% a, 0% b, 50% c, 50% d) → entropy = 1.00

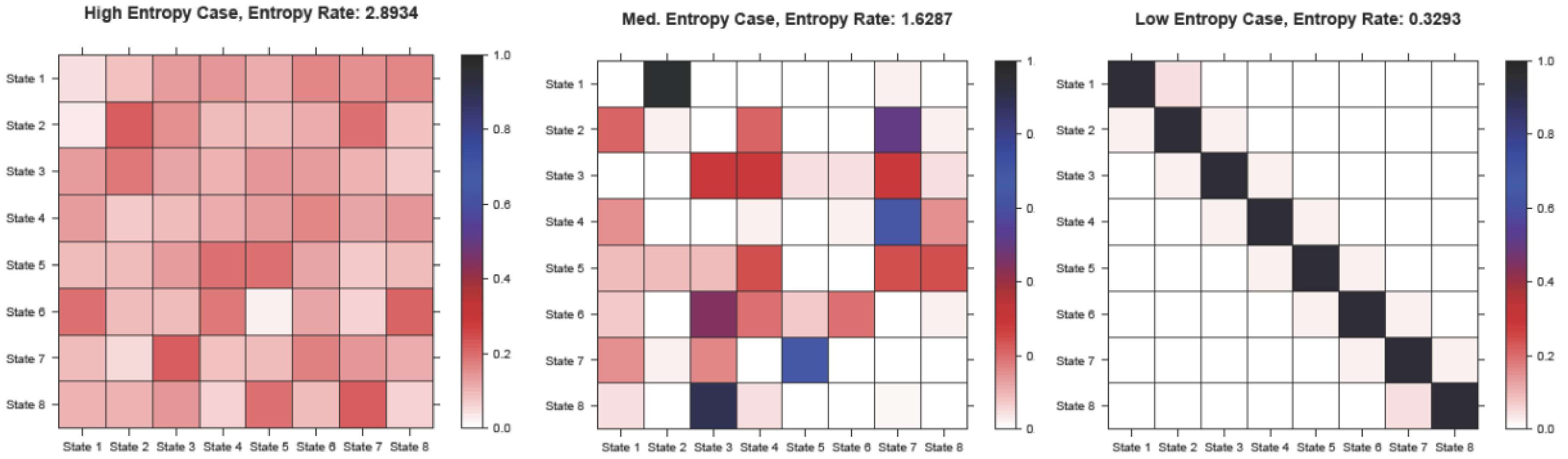
Entropy

- Entropy of a random variable or a distribution
  $H = - \Sigma p_i \log p_i$ where p_i is the probability of seeing outcome i
- In our case, we are interested in the entropy of a sequence of behaviors or entropy of a random process $X_0, X_1, X_2, \ldots X_T$ where X_t is the t-th observed behavior
- At least two approaches ….
  1. Build a probability model for X_t conditional on previous observations (e.g., a 1^{st}-order Markov chain model) and compute entropy as limiting value of entropy of the conditional distribution
     - Computation straightforward
     - May be sensitive to choice of model

  2. Think about a coding or data compression approach to describing the sequence of behaviors. Theoretical result relates entropy to the compression rate.
     - Computation more complex
     - Does not require a model assumption

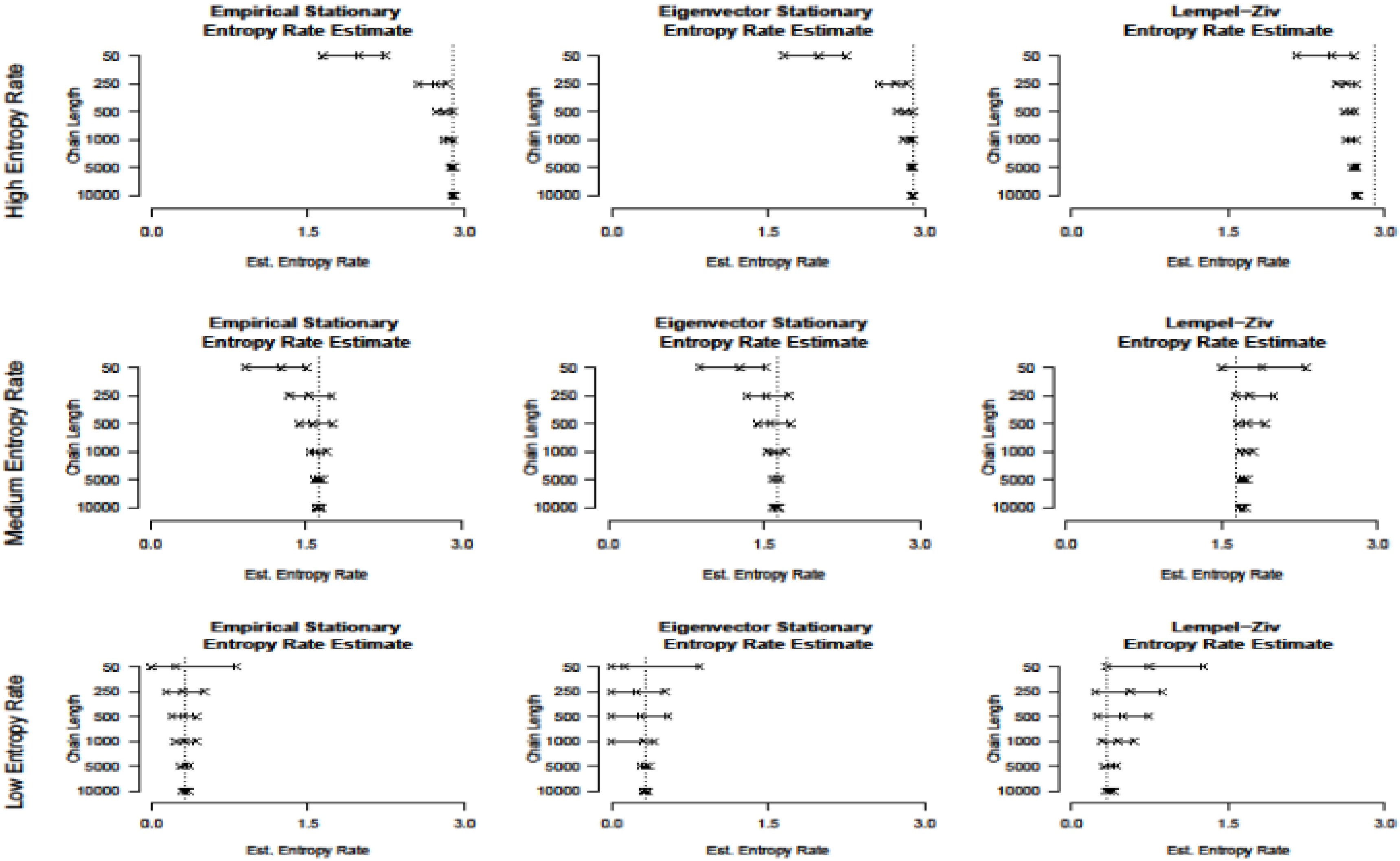
Entropy

- Examples of entropy for 1-st order Markov chain behavior



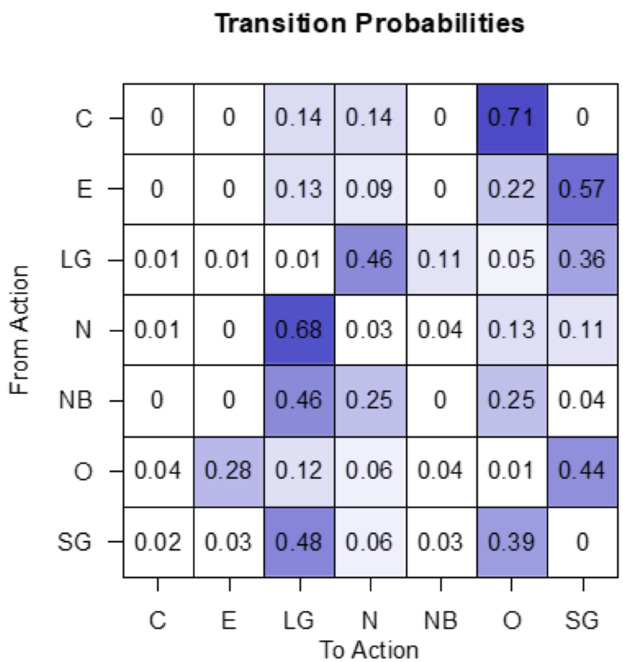
Entropy

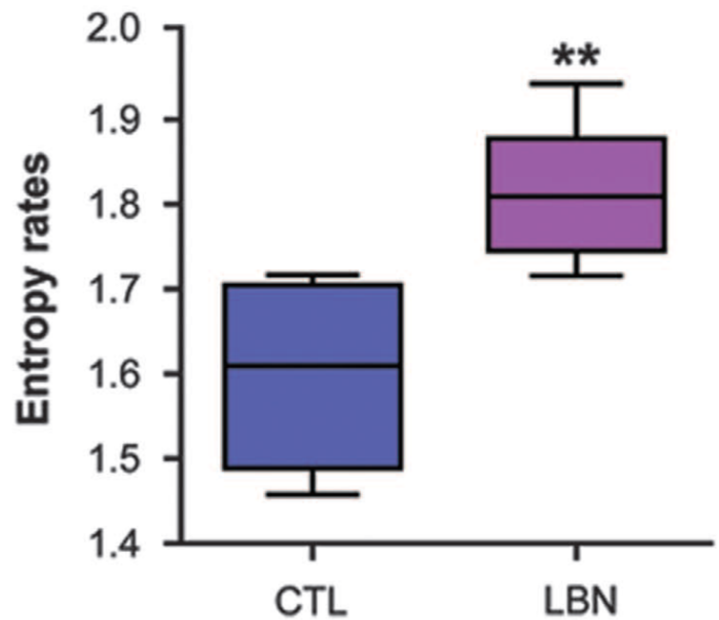
- Simulation study of different approaches to estimating the entropy rate (Vegetabile et al., 2019)
- Rows= Different entropy levels; Columns= Different estimation methods



Assessing predictability - rodents

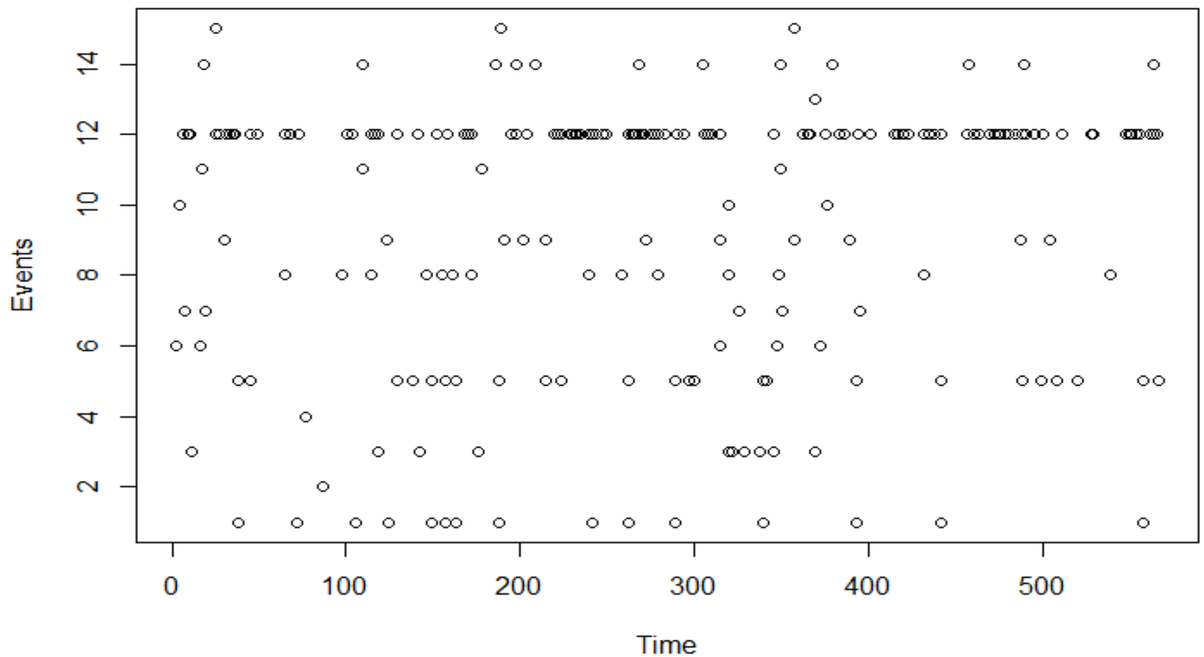
- Rodents





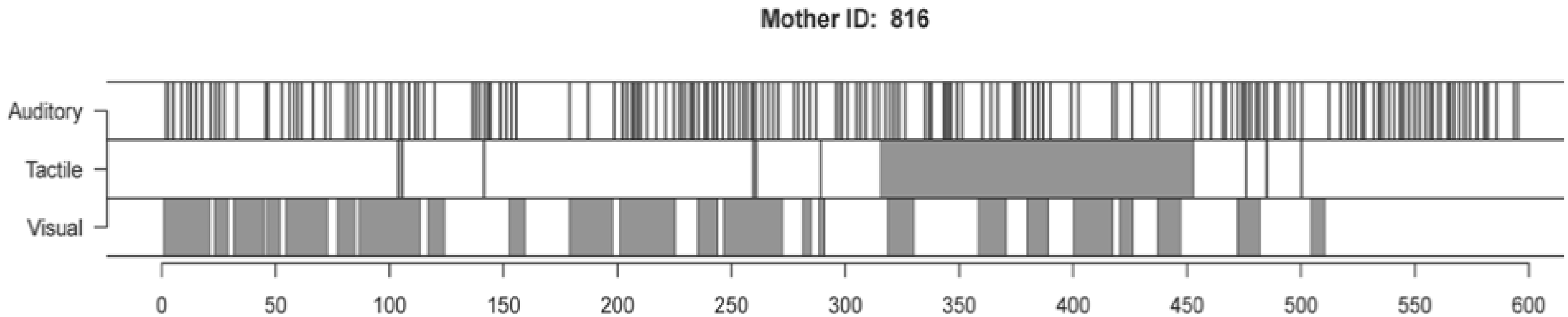
Assessing predictability - humans

Initial attempt based on a large number of recorded behaviors



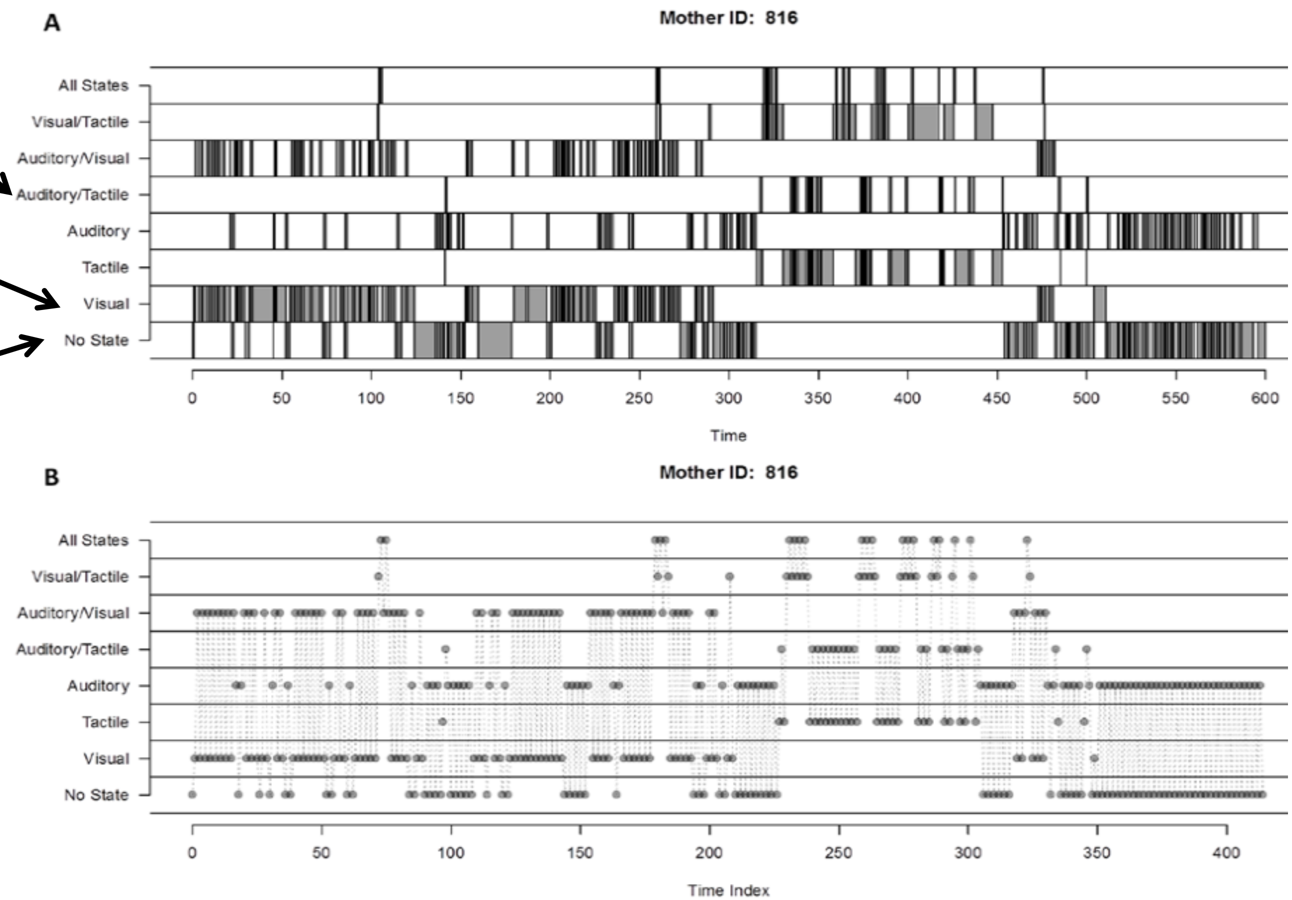
Assessing predictability - humans

For one mother-child pair, record of the three sensory channels being engaged/not



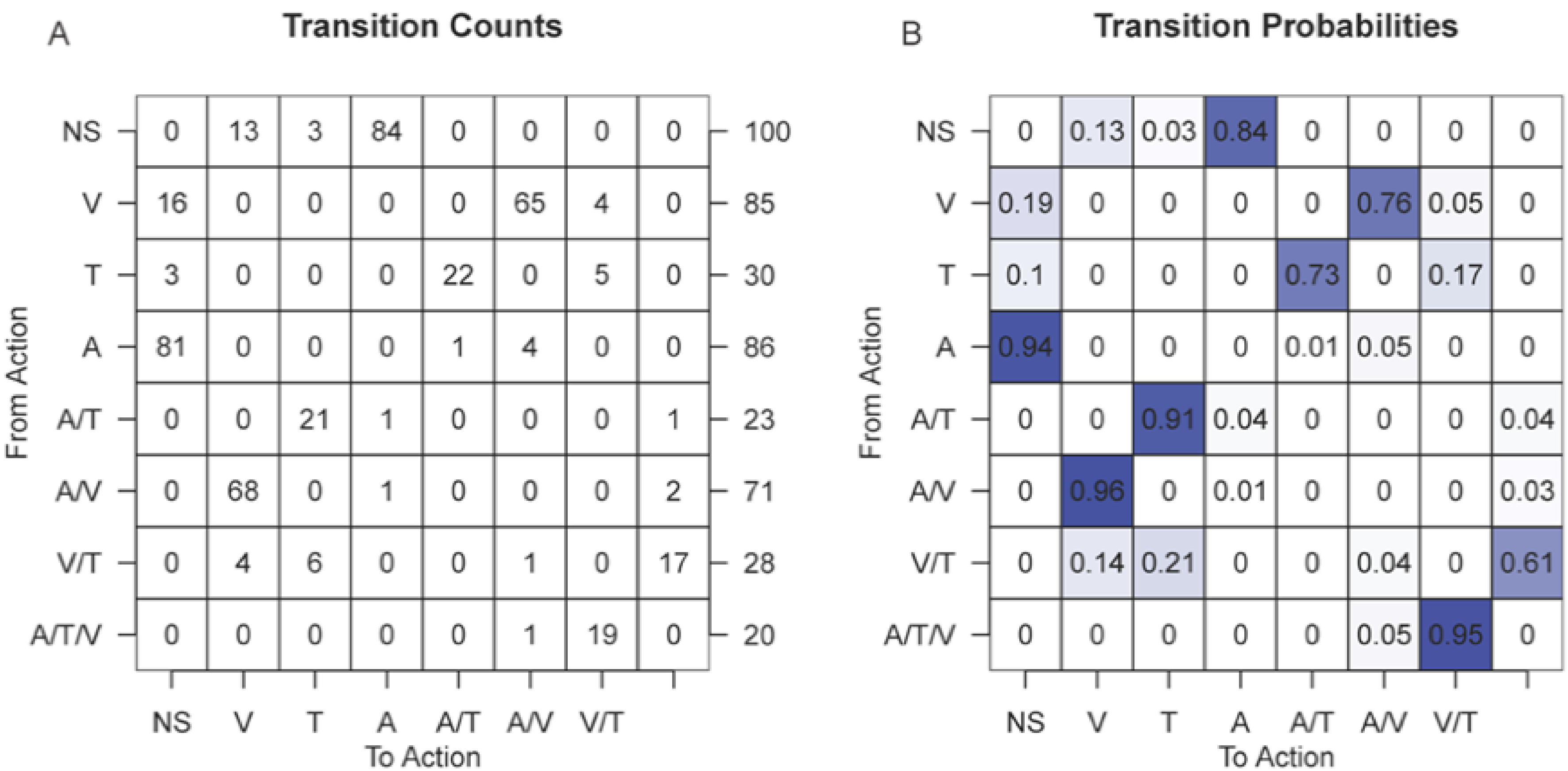
Assessing predictability of mom

- Define behavioral states with respect to which sensory channels are in use
- For example:
  - Mother speaking and touching
  - Mother looking at child while child attends
  - Mother not interacting with child



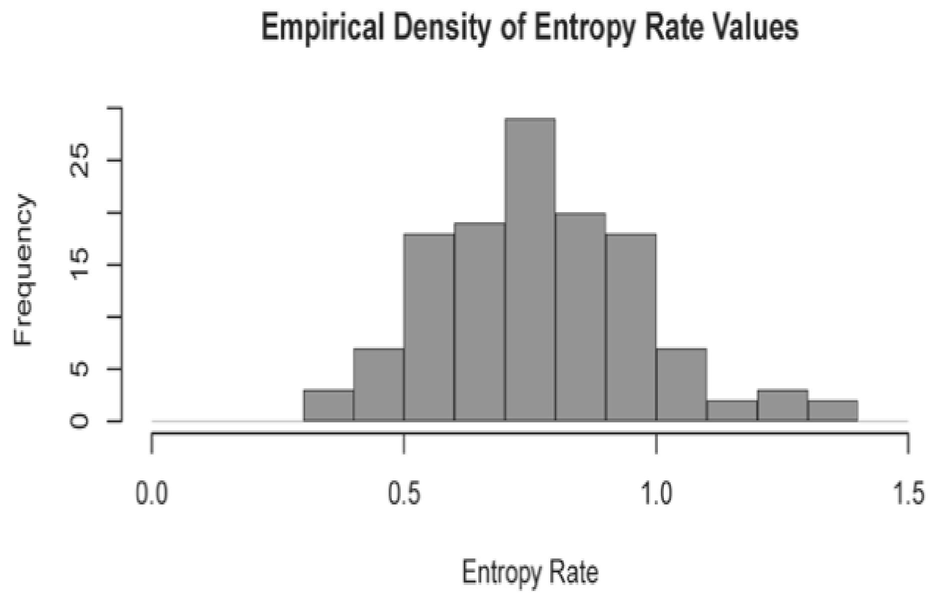
Assessing predictability of mom

- Summarize by looking at transitions between states (counts, probabilities)



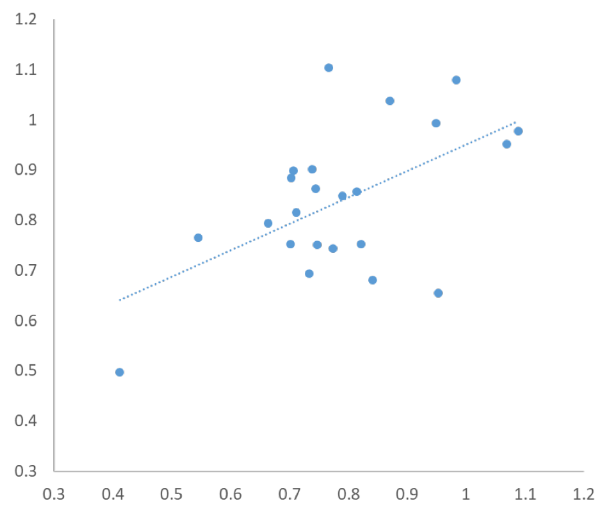
Assessing predictability of mom

- Computed the entropy rate for each mother assuming first-order Markov process
- Also considered alternative models



Reliability of the entropy measure

- Entropy computed from first five minutes of the 10-minute play session is correlated with entropy computed from the second five minutes (r = 0.5)
- Entropy computed from mother-child dyad at 6 months is correlated with mother-child dyad at 12 months (r=0.4)
- Mothers with two children in the studies have similar entropy (r = 0.55)



- Now have entropy from 2^{nd} cohort … similar results and relationships

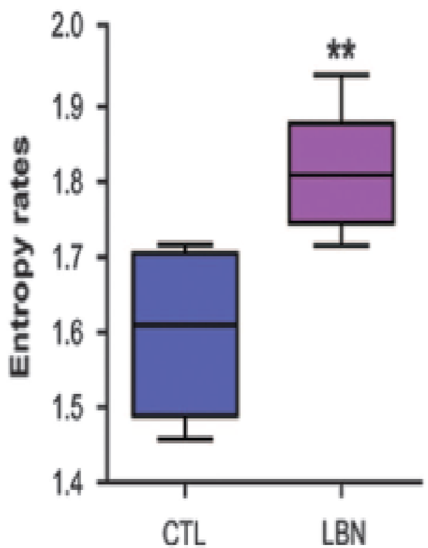
UCI Department of Statistics
Donald Bren School of Information & Computer Sciences

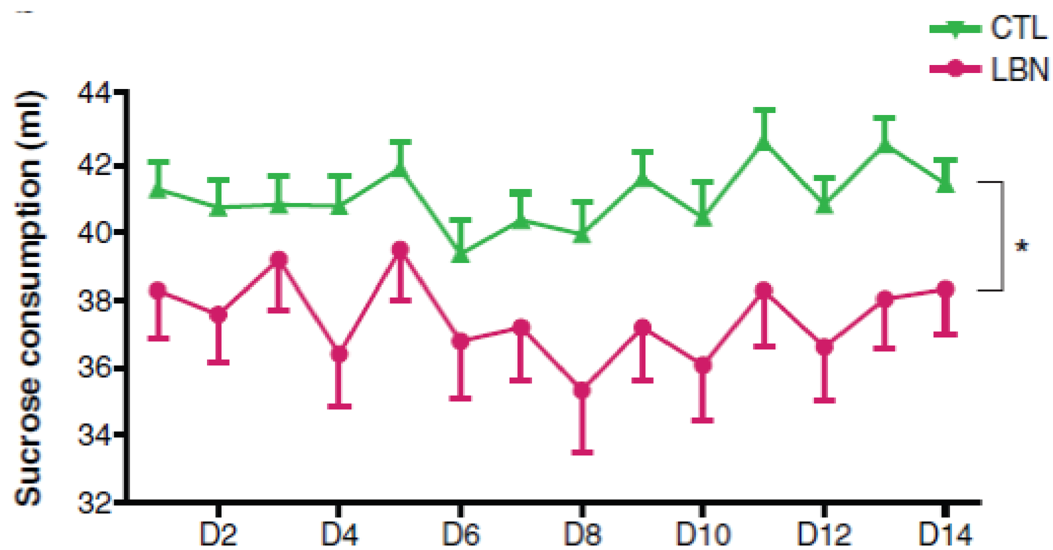
Other measures of unpredictability

- Household chaos
- Maternal mood
  - Mood scales completed by mothers at 15, 20, 25, 30, 35 weeks prenatally and several time points postnatally
  - Scales measured: depressive symptoms, state anxiety, pregnancy-specific anxiety, perceived stress
  - Alternative approach – examined consistency / randomness of item responses within each scale at a given time point
  - Averaged this measure across scales and time points -> maternal mood entropy

Effects of unpredictability in rodents

- We take the same steps to measure entropy in rodent mothers
- Mothers in LBN cages have much higher entropy (i.e., are less predictable)
- Their offspring are vulnerable to emotional/cognitive issues
- Offspring also demonstrate much reduced interest in pleasurable activities (e.g., sugar consumption) → anhedonia
- Other signs of anhedonia as well

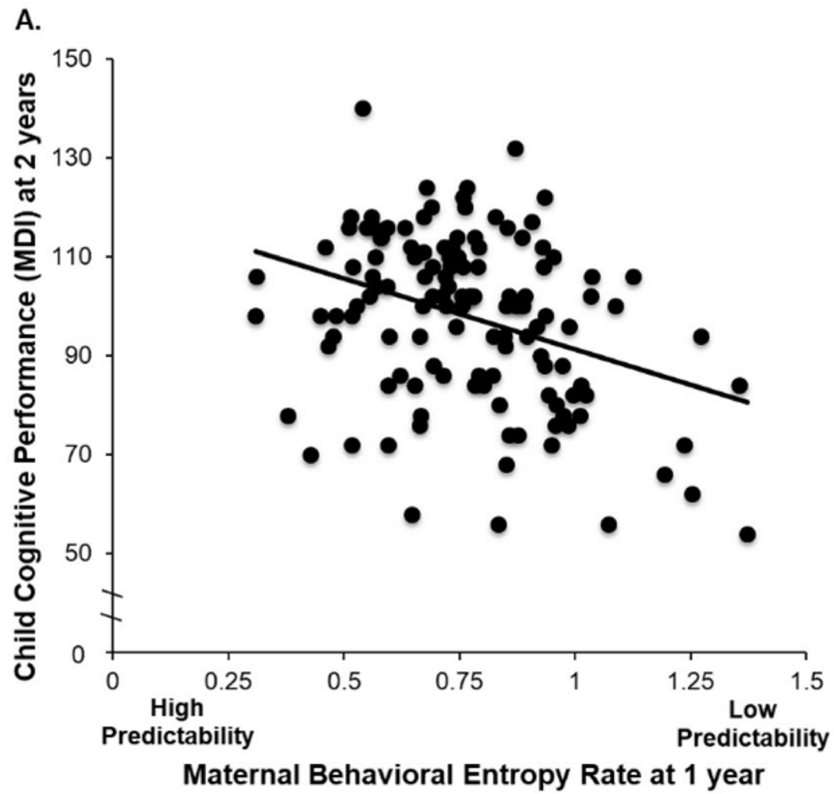




UCI Department of Statistics
Donald Bren School of Information & Computer Sciences

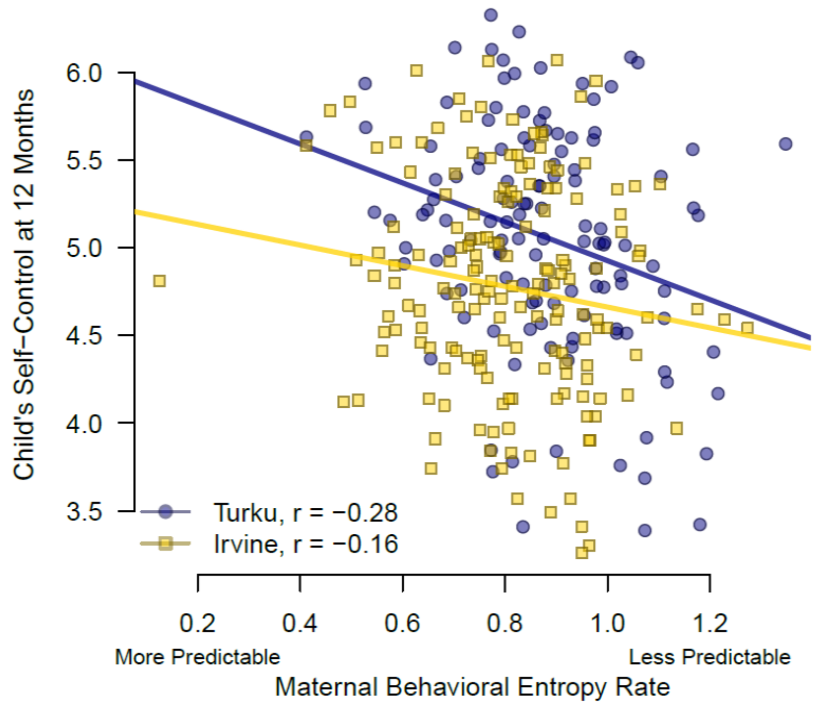
Effects of unpredictability in humans?

- Maternal behavioral entropy at 1 yr visit is associated with child's cognitive performance at 2 years (and later in life)



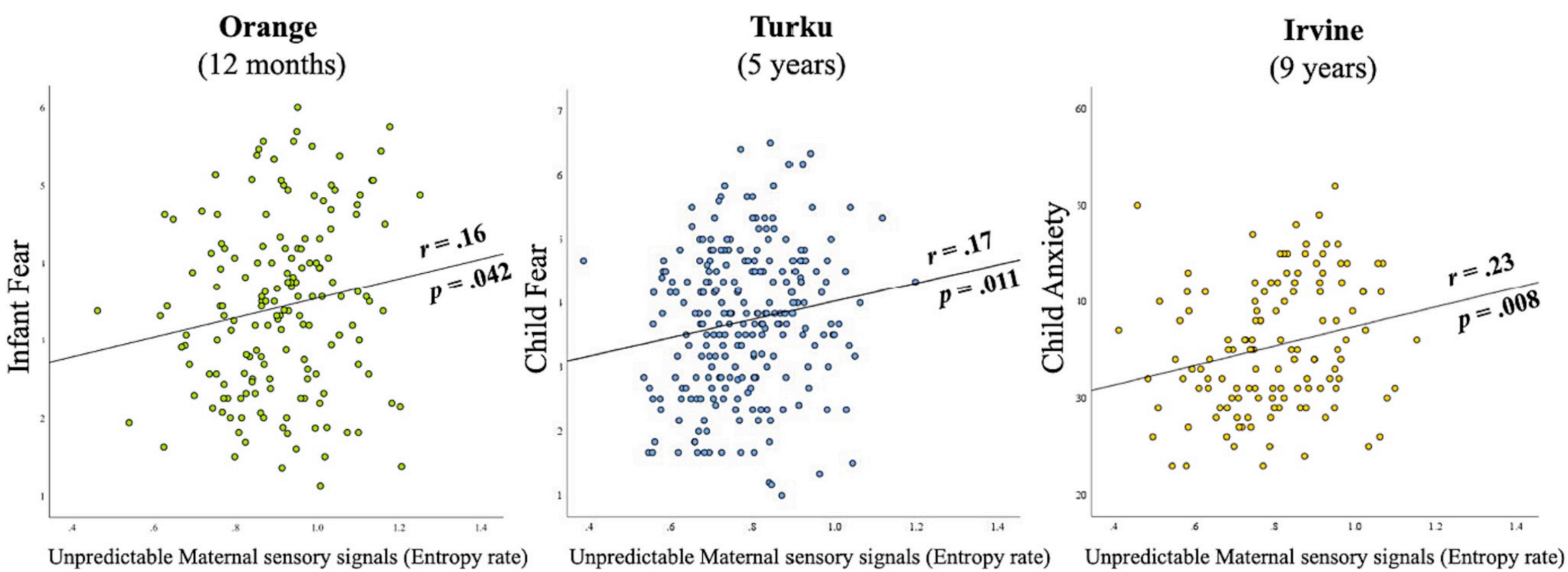
Effects of unpredictability in humans?

- Maternal behavioral entropy is associated with child self-control in two very different cohorts (Irvine, CA and Turku, Finland)



Effects of unpredictability in humans?

- Maternal behavioral entropy is associated with child outcomes across different ages
- Here we show patterns for internalizing behaviors (fear/anxiety)
- True also for effortful control across ages



UCI Department of Statistics
Donald Bren School of Information & Computer Sciences

Translation beyond childhood

- New project with marine veterans for the last 5 years
  - Relate early life environment to anhedonia
  - Relate early life environment and anhedonia with vulnerability to PTSD and other mental illness
- Challenge –
  How do we assess early life environment for these young adults?

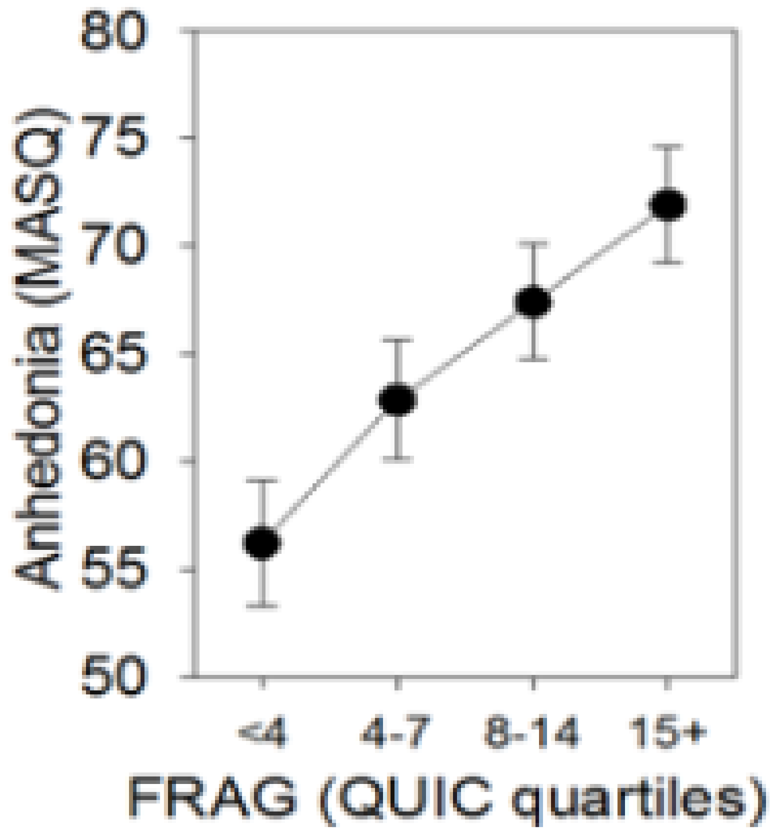
A self-report measure of childhood unpredictability

- Questionnaire of Unpredictability in Childhood (QUIC) (Glynn et al. 2018)
- 38 items organized in 5 subscales (parental predictability, parental environment, parental monitoring, physical environment, safety) Examples:
  - There were often people coming and going in my house that I did not expect to be there
  - I experienced changes in my custody arrangement
  - I often wondered whether or not one of my parents would come home at the end of the day
- Tested on 3 cohorts
  - Adult females (mothers of adolescents) (n=116)
  - Adult males (marines) (n=95)
  - Adolescents (n=175)

A self-report measure of childhood unpredictability

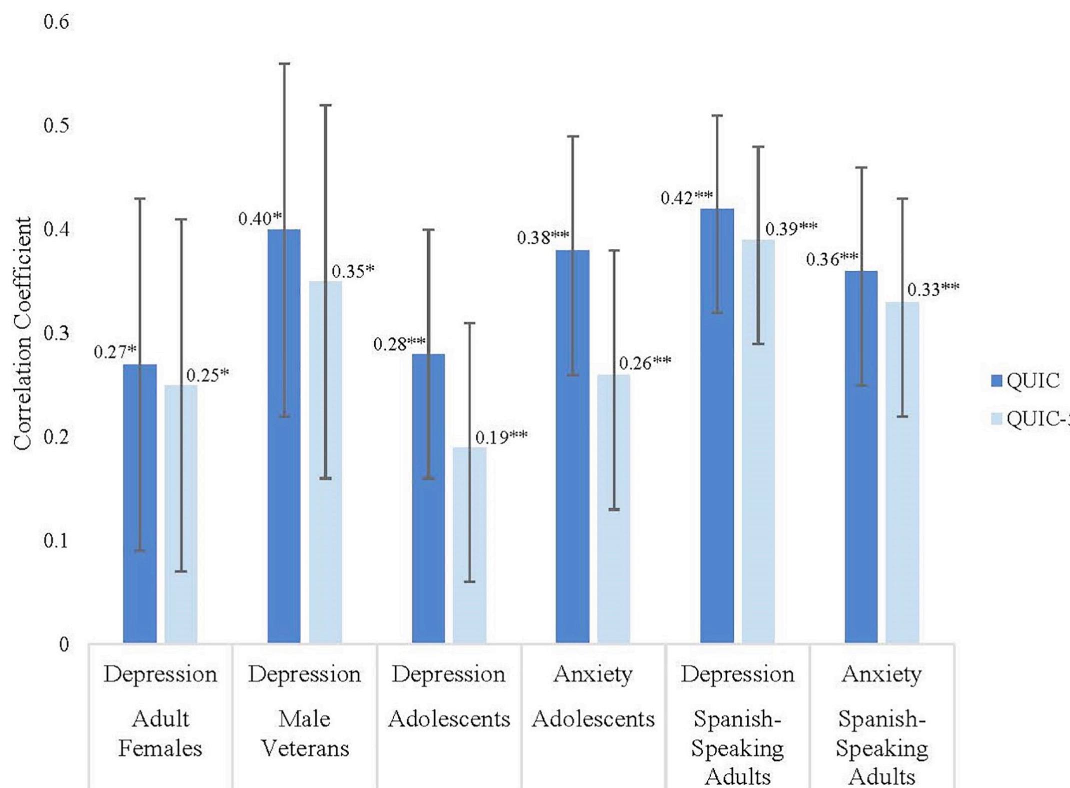
- Good psychometric properties
  - Test-retest reliability (r=.92)
  - Internal consistency (alpha=.84 - .92)
  - Adolescent data validated by prospective reports for some items (e.g., moving)
  - QUIC correlated with other measures of traumatic/stressful life events
  - **Adolescent QUIC scores correlate with maternal behavioral entropy (r=.23)**
  - QUIC predicts mental health risk (anhedonia, depression, anxiety)

Effect of Unpredictability in Marines



A self-report measure of childhood unpredictability for clinical use

- The QUIC has proven to be a useful scale for research
- Adopted by other labs
- But of limited use for clinical application
- We have developed a 5-item version (more on this later)



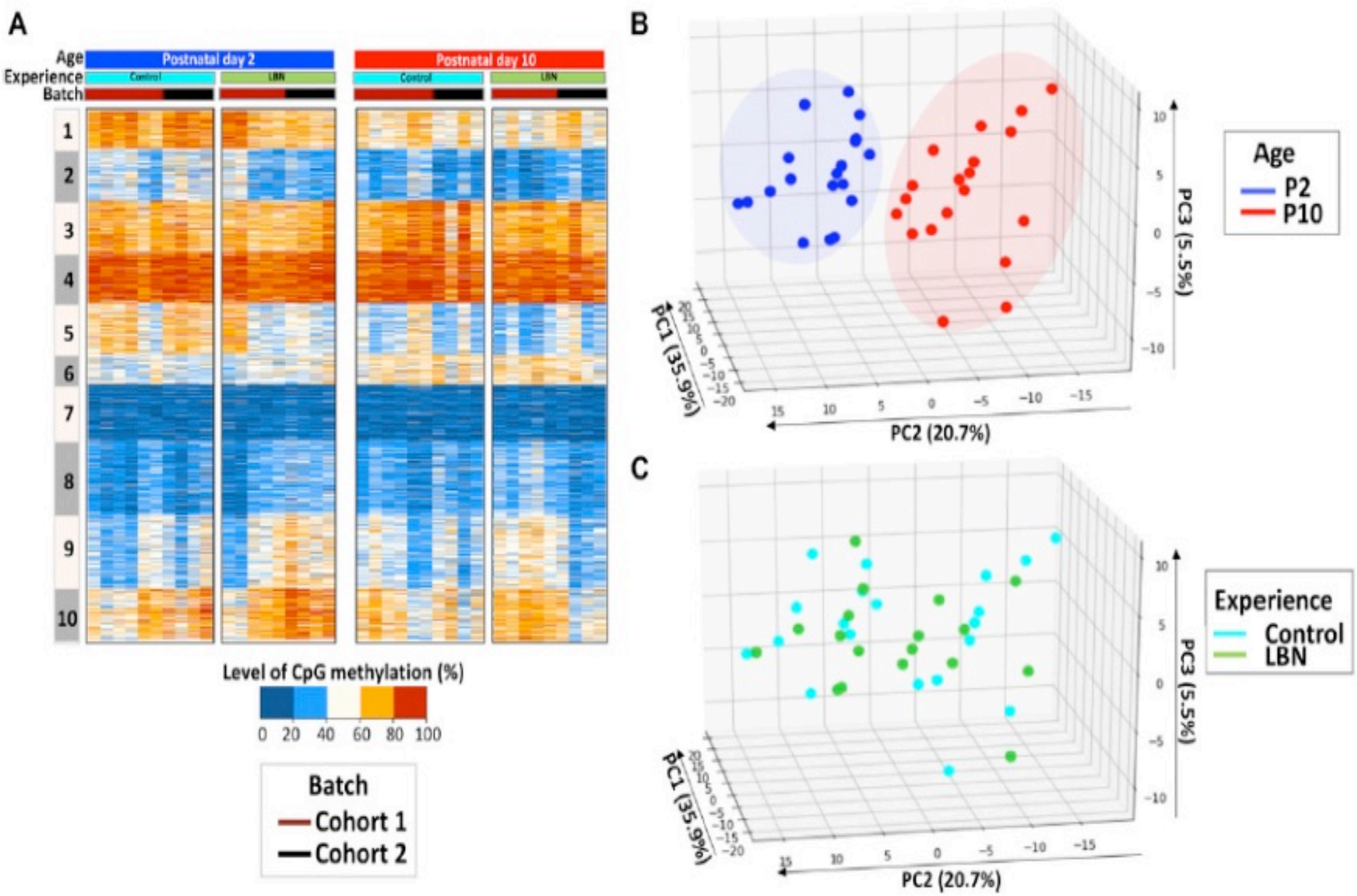
Exploring mechanism - methylomics

- Environment can effect expression of genes
- One mechanism is DNA methylation
- CpG sites (cytosine (C) follow by guanine (G)) can be methylated (a methyl group is added)
- This can change the expression of the associated gene
- We explore whether methylomics may be a way in which early-life adversity (unpredictability) leaves a “mark”

- One challenge that has been observed in methylomic analyses is that there is considerable inter-individual variation in methylation levels

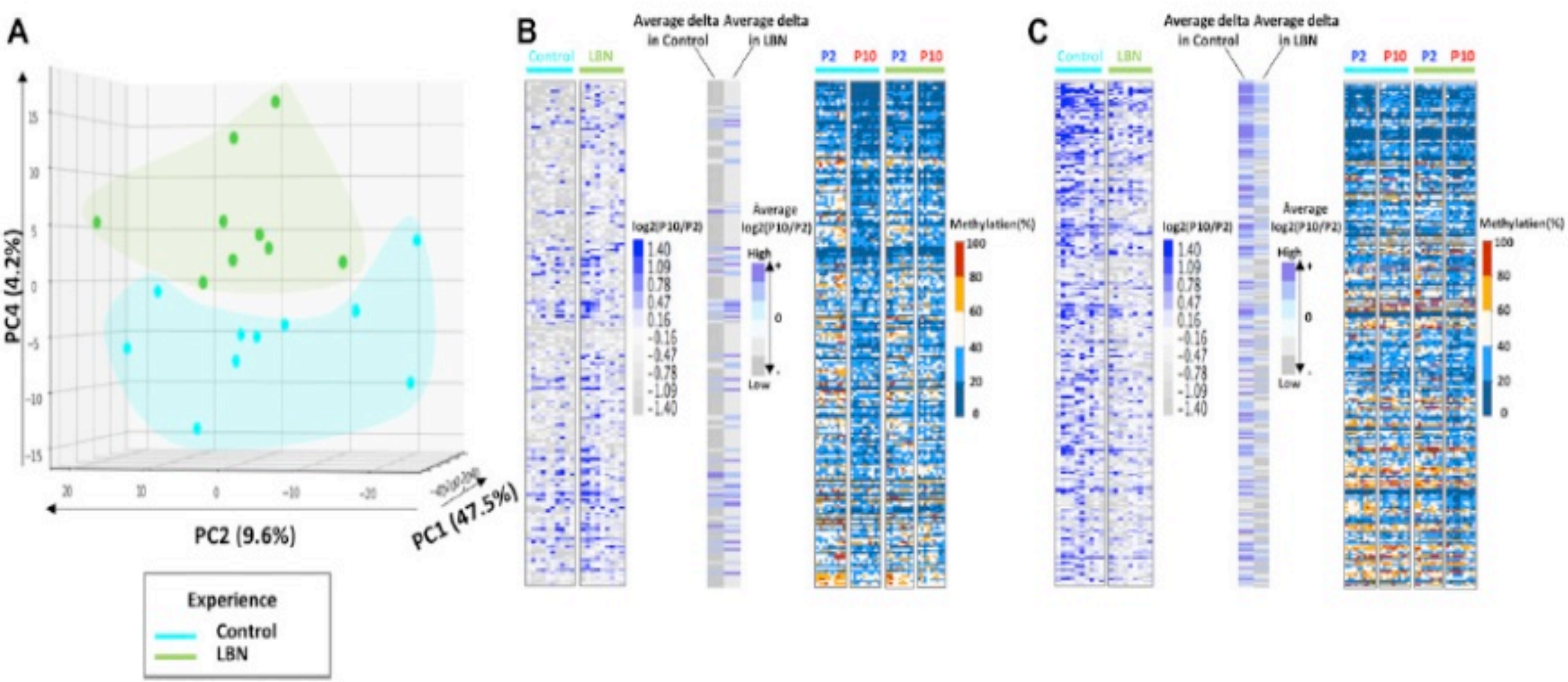
Methylomics in rodents

- An initial analysis of rodent data incorporating day 2 and day 10 samples distinguishes age but not CTL/LBN



Methylomics in rodents

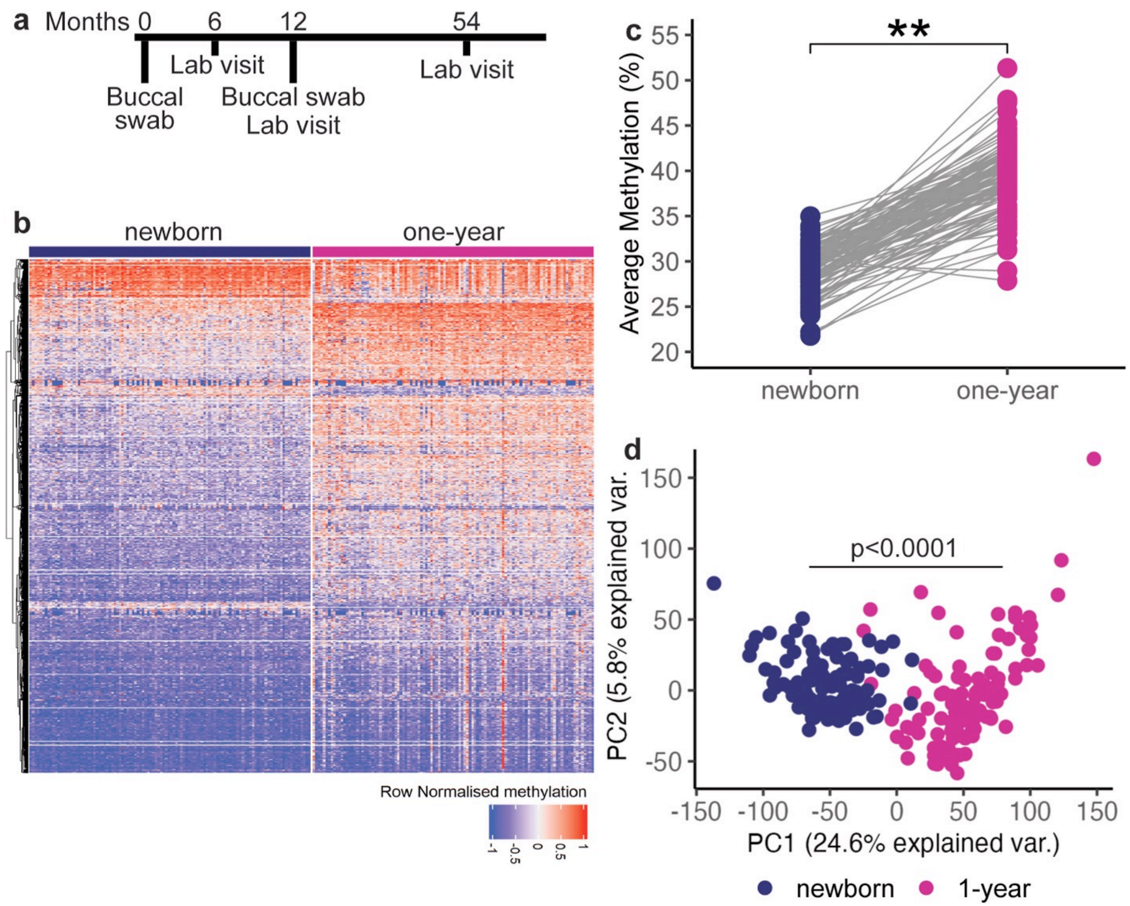
- Considerable inter-individual variation
- An alternative analysis focuses on intra-individual change in methylation (log (P10 / p2))



UCI Department of Statistics
Donald Bren School of Information & Computer Sciences

Methylomics in humans

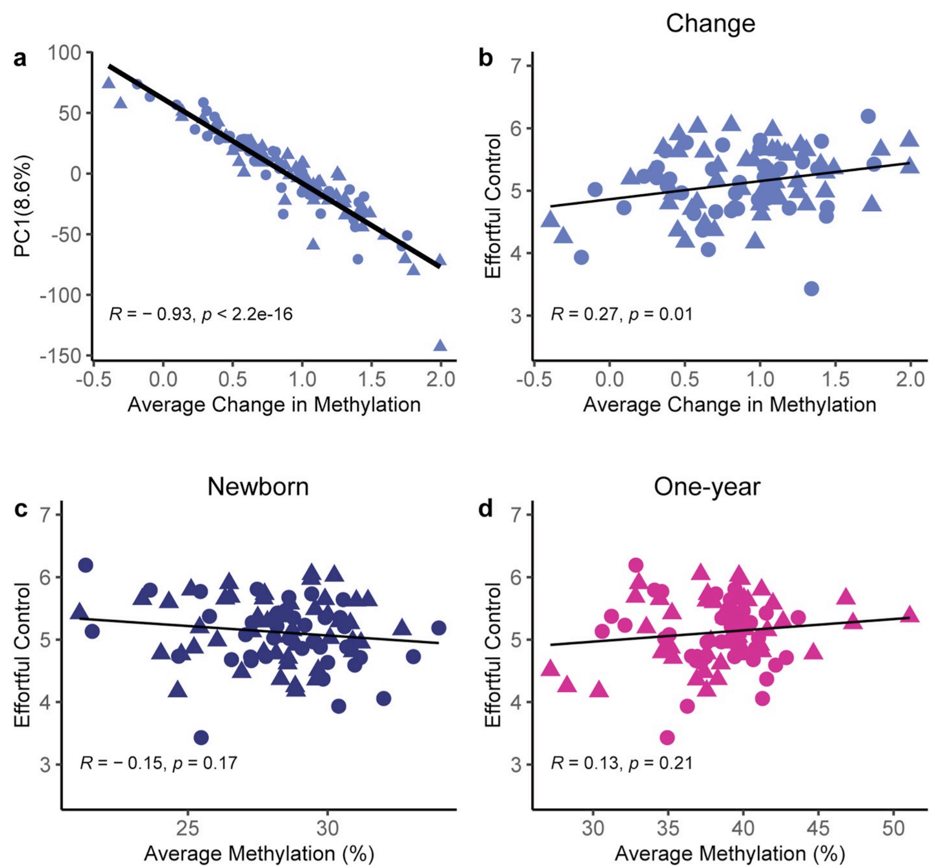
- An initial analysis of human data incorporating newborn and 12-month samples distinguishes age (which is not very interesting)



UCI Department of Statistics
Donald Bren School of Information & Computer Sciences

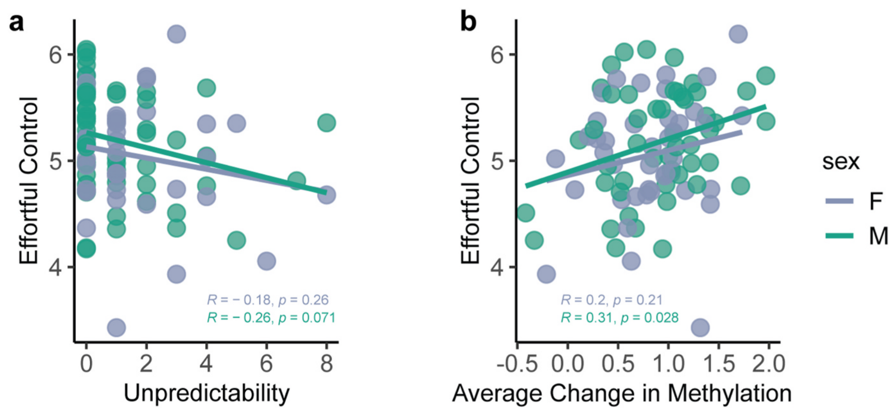
Methylomics in humans

- Apply the same approach that was useful in the rodents
- Consider the delta/change in methylation over the two time points
- Plot a shows that the first principal component of the delta methylation scores is essential average change
- So we use average change
- Plot b shows that average change in methylation is associated with effortful control in children
- Plots c,d show that individual methylation measures are not



Methylomics in humans

- Established literature (earlier in the talk) that unpredictability is associated with outcomes in humans (figure a below)
- Previous slide shows methylomics is associated with outcomes in humans (figure b repeated here)



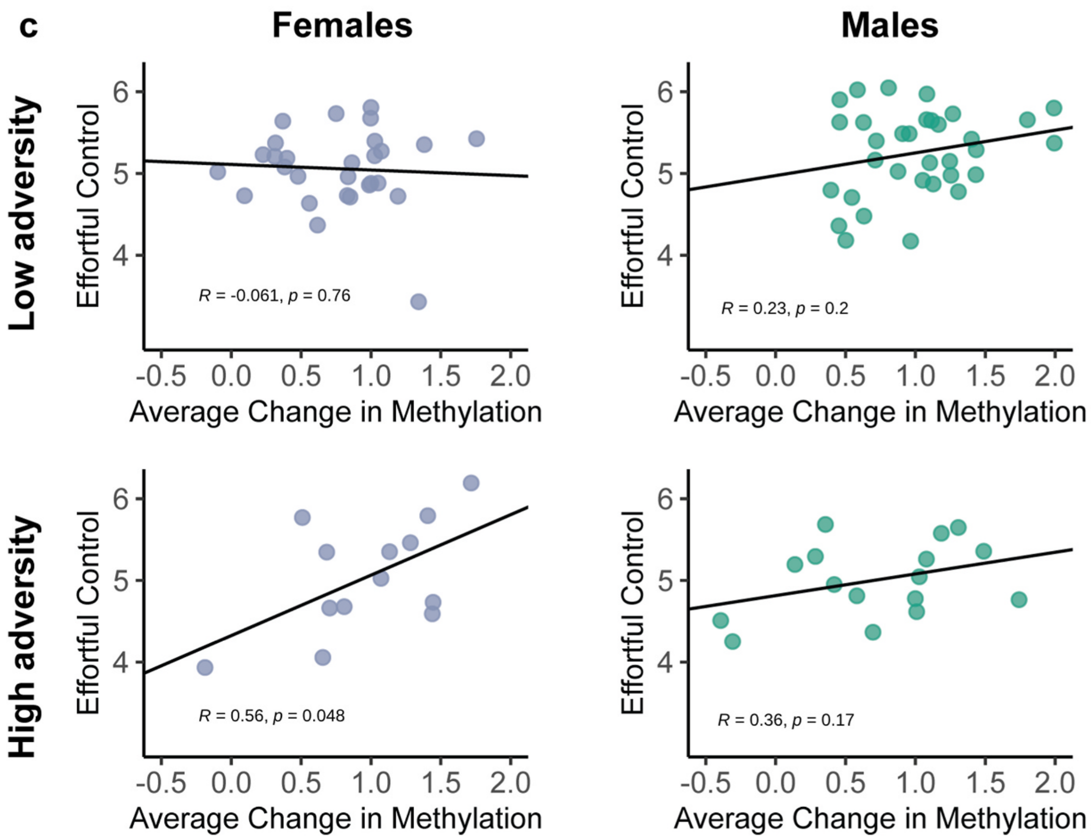
- Animal studies found that methylomics could distinguish control and LBN rodents
- For humans, average change in methylation and unpredictability are not directly related ($r = -0.07$)

Methylomics in humans

- Recall that for animal studies, methylomics could distinguish control and LBN rodents (unpredictability), and unpredictability was associated with child/adolescent outcomes
  - Other studies (not discussed here) show that unpredictability appears to create a "vulnerability" to future adversity
- For humans:
  - Average change in methylation and unpredictability are only weakly related ($r = -0.07$)
  - We explore whether unpredictability and methylomics may interact in their impacts on children

Methylomics in humans

- Some preliminary evidence that unpredictability may impact the relationship of methylation and outcomes in a sex-dependent way, i.e., an interaction



Conte Center Next Steps

- Next steps
  - Funding from a CA Precision Medicine Award to investigate effects of early-life unpredictability in population
    - Partnerships with clinics
    - Data on thousands of children (QUIC5 and outcomes)
  - Methylomics
    - Validation sample from the Precision Medicine study
    - Partnership with Finland collaborators

Summary

- Novel early-life experience (maternal unpredictability) developed in a rodent model
- Interrogated through a variety of data analysis approaches
  - Application of entropy (across species) to characterize unpredictability
  - Standard statistical analysis (correlation/regression) associating unpredictability with a range of outcomes
  - High-dimensional AI/ML approaches to explore genetic markers or impact of unpredictability
- Importance of collaboration and team science
- Contact: sternh@uci.edu

UCI Department of Statistics
Donald Bren School of Information & Computer Sciences