PSYCO 452

Week 10: Exploring Distributed Representations

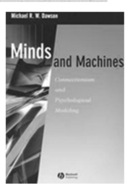
- •Algorithms From Network Interpretations
- •Chord Classification
- •Distributed Representation Examples
- •Translating Classical Theories Into Connectionist Networks

Course Structure

When	What
Weeks 1, 2, 3	Connectionist Building Blocks
Weeks 4, 5, 6	Case Studies of Connectionism
Week 7	Midterm Exam
Weeks 8, 9, 10	**Interpreting Connectionist Networks**
Weeks 11, 12	Deep Learning Basics
Week 13	Final Exam

Chapter 6 Discussion

- Questions?
- Important Terms
  - Synthetic psychology
  - Embodied cognitive science
  - Synthesis
  - Emergence
  - Analysis
  - SEA
  - Thoughtless walker
  - Recognizable and recurring patterns
  - Rule-governed system
  - Dynamic system
  - Adaptive system

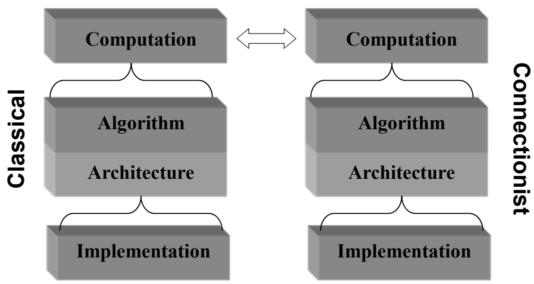


Tri-Level Consideration

- Classical and connectionist cognitive science are frequently portrayed as being antagonistic opposites
- However, my own work is interested in exploring similarities between the two approaches
- This is done in the context of the tri-level hypothesis

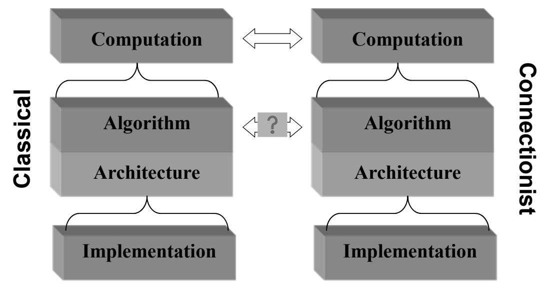
Computational Equivalence

- Many different kinds of proofs exist suggesting that PDP networks are equivalent to UTMs



Algorithmic Equivalence?

- What kinds of algorithms do networks execute?
- Can they be related to classical algorithms?



PDP Models Are Hard To Understand

- **Problem: researchers rarely describe network algorithms, because network interpretation is not an easy task**
- "If the purpose of simulation modeling is to clarify existing theoretical constructs, then connectionism looks like exactly the wrong way to go. Connectionist models do not clarify ideas, they obscure them" (Seidenberg, 1993)

Mark Seidenberg

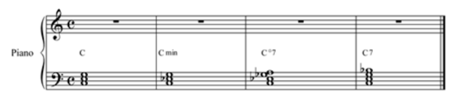
Synthesis, Emergence, Analysis

- However, if you go to the trouble of peering into networks, you can be rewarded
- My students and I have spent a great deal of time interpreting PDP networks
- Synthesis
  - Build a network
- Analysis
  - Interpret its internal structure
- Emergence
  - Learn surprises about the phenomena by discovering network properties

Case study from music: Distributed representations based on 'strange circles'

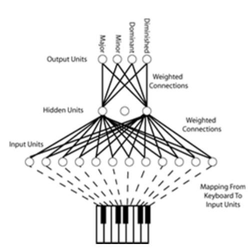
Chord Classification Problem

- One important task in music theory training and piano technical training is chord identification
- Example: listen to a chord
  - What general type of chord is it?
  - Independent of key
  - Independent of inversion



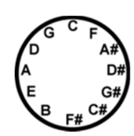
The Pitch Class Network

- 4 output value units
  - Major chord
  - Minor chord
  - Dominant chord
  - Diminished chord
- 3 hidden value units
- 12 input units
  - Piano keyboard
  - One octave
  - Starting note is A
- 48 training patterns
- Dawson/Schopflocher rule
  - Learning rate of 0.005
  - Weight start ±0.10
  - Biases start at 0.00
- Converged after 3964 epochs

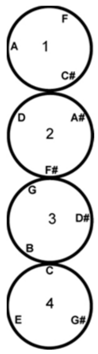


Network Analysis

- "Gee Whiz connectionism" is no more
- To find surprises, or emergent properties, you have to analyze internal properties first!
- We focused on the relation between connection weights and note names
- We found a set of equivalence classes similar to the 'circle of fifths', but based on other intervals between notes

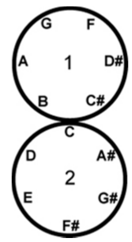


Circles Of Major 3rds



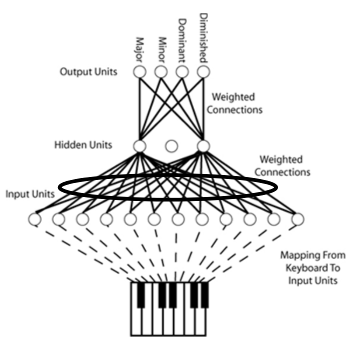
- One can create four different circles of major 3rds
- Each circle has three notes
- As you move from one note in the circle to the next, you cover an interval of a major 3rd (4 semitones)

Circles Of Major 2nds

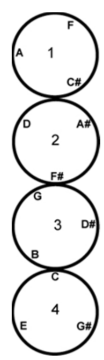


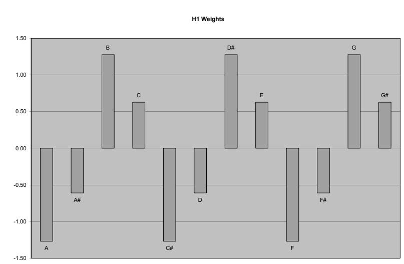
- One can create two different circles of major 2nds
- Each circle has six notes
- As you move from one note in the circle to the next, you cover an interval of a major 2nd (2 semitones)

Examining First Layer Connections

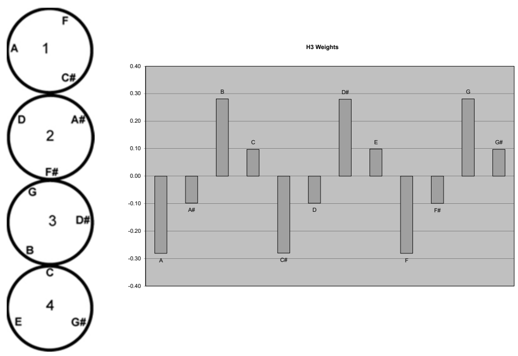


H1 Weights And Circles Of Major 3rds

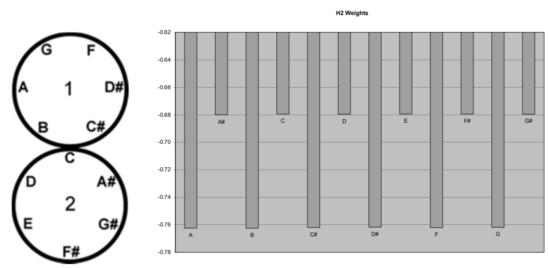




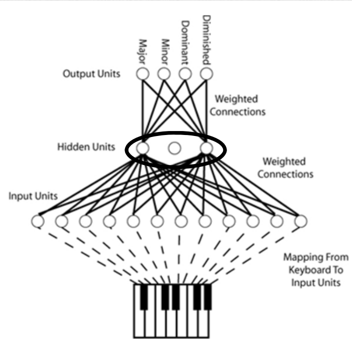
H3 Weights And Circles Of Major 3rds



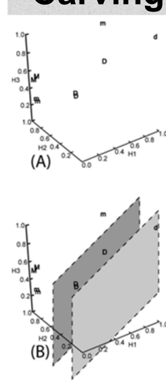
H1 Weights And Circles Of Major 2nds



Examining Second Layer Processing



Carving Hidden Unit Space



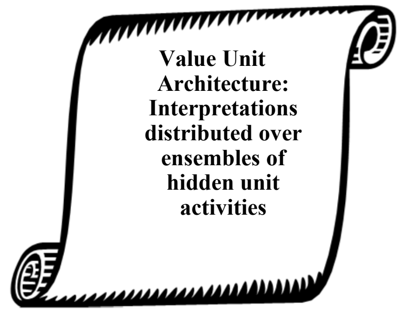
- The chords are arranged in a 3D hidden unit space – coarse coding based on the strange circles
- Output value units "carve" two parallel planes through this space
- Each unit can carve the space to separate one chord type from all others

Implications

- Our network outperformed earlier networks of Laden and Keefe
- Interpretation of the network revealed an unusual set of equivalence classes of notes
- Results in a new understanding of musical regularities, and makes some predictions that can be explored by studying human listeners

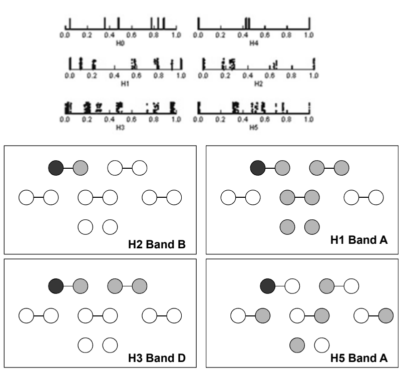
Bunny Laden

Douglas Keefe



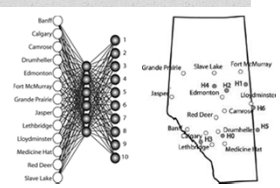
Dawson & Piercey (2001)

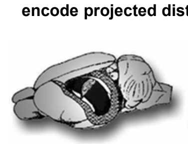
- Hinton's kinship problem
- "Who is James' father?" "Andrew"
- 6 families, 52 queries per family, 312 patterns
- 21 inputs, 6 hidden, 9 output
- Local bands uninterpretable
- Intersection of bands results in clean coarse coding interpretation

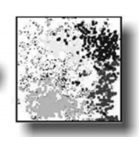


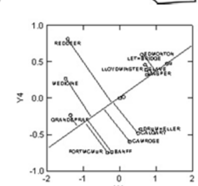
Dawson, Boechler & Valsangkar-Smyth (2000)

- Network trained to rate distances between Alberta cities
- Is there an internal spatial representation?
- Hidden units analogous to place cells in rat hippocampus
- Connection weights metric – encode projected distances



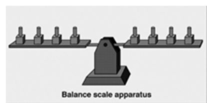


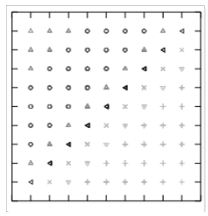




Dawson & Zimmerman (2003)

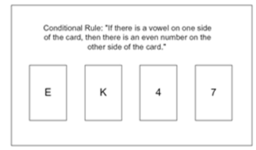
- Network trained to make ideal responses to Piaget's balance scale task
- Four hidden units use course coding to determine whether balance scale will tip left, tip right, or stay balanced
- Interpretation of network led to new additive rule for defining behaviour of balance scale
- Interpretation of network led to a new classification of problems based on a novel 2D pattern space

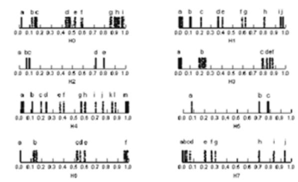


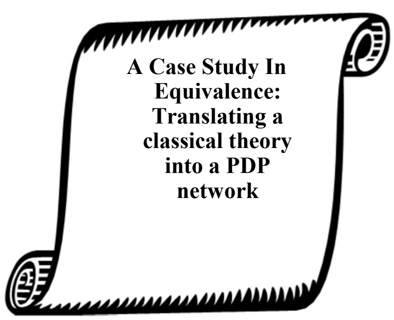


Leighton & Dawson (2001)

- Series of networks trained to give different kinds of responses to Wason Card Selection Task
- All hidden units produce bands
- Bands support an inductive set of rules for solving this task, instead of a more traditional deductive theory
- Interpretations also were used to assess difficulty of different kinds of responses







Theory Translation

- If two theories are really qualitatively different, then you can't translate one into the other
- Is this true for symbolic and connectionist theories?

Thomas S. Kuhn

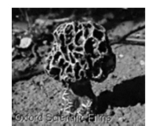
The Mushroom Problem

- Problem: determine whether a mushroom is poisonous or not
- Consider 8124 different mushrooms
- Each mushroom is described using values on 21 different features

Theory 1 (Classical)

- What is the mushroom's odor?
  - If almond or anise then edible
  - If another definite odor then poisonous
  - If no odor then go to next step
- What is the spore print color?
  - If white then go to next step
  - If green or purple then poisonous
  - If some other color then edible
- What is the gill size of the mushroom?
  - If broad then edible
  - If narrow then go to next step
- Examine the stalk surface above the mushroom's ring
  - If fibrous then edible
  - If silky or scaly then poisonous
  - If smooth then go to next step
- Does the mushroom have bruises?
  - If not, then edible
  - If it does, then poisonous

Deadly



Tasty!

Decision Tree To Production System

Odor: Almond or Anise

Odor: Creosote or Fishy or Foul or Musty or Pungent or Spicy

Odor: None
Spore Print: Black or Brown or Buff or Chocolate or Orange or Yellow

Odor: None
Spore Print: Green or Purple

Odor: None
Spore Print: White
Gill Size: Broad

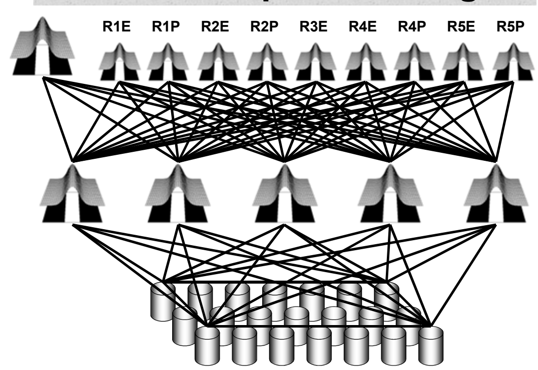
Odor: None
Spore Print: White
Gill Size: Narrow
Stalk Surface Above Ring: Fibrous

Odor: None
Spore Print: White
Gill Size: Narrow
Stalk Surface Above Ring: Silky or Scaly

Odor: None
Spore Print: White
Gill Size: Narrow
Stalk Surface Above Ring: Smooth
Bruises: No

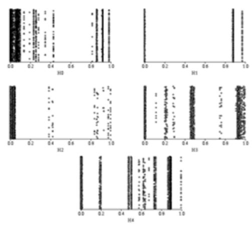
Odor: None
Spore Print: White
Gill Size: Narrow
Stalk Surface Above Ring: Smooth
Bruises: Yes

Extra Output Learning



Hidden Unit Banding

- The hidden units of this network demonstrate a high degree of banding
- Can be locally interpreted
- Distributions over hidden units can also be interpreted

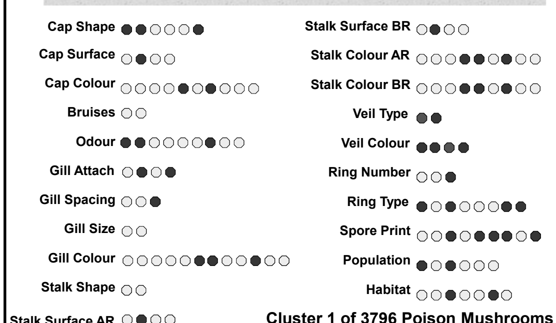


Theory 2 (Connectionist)

CLUSTER	POISONOUS	EDIBLE
1	3796	0
2	0	704
3	0	96
4	0	528
5	40	0
6	72	0
7	0	12
8	0	12
9	0	2832
10	8	0
11	0	12
12	0	12

Each cluster is "pure" in terms of network's main response

Definite Features



Cluster 1 of 3796 Poison Mushrooms

Each cluster is laden with definite features

Clusters Map Onto Productions!

Odor: Almond or Anise — C1

Odor: Creosote or Fishy or Foul or Musty or Pungent or Spicy — C2, C3

Odor: None — C9
Spore Print: Black or Brown or Buff or Chocolate or Orange or Yellow

Odor: None — C6
Spore Print: Green or Purple

Odor: None — C4
Spore Print: White
Gill Size: Broad

Odor: None — C8, C12
Spore Print: White
Gill Size: Narrow
Stalk Surface Above Ring: Fibrous

Odor: None — C5
Spore Print: White
Gill Size: Narrow
Stalk Surface Above Ring: Silky or Scaly

Odor: None — C10
Spore Print: White
Gill Size: Narrow
Stalk Surface Above Ring: Smooth
Bruises: No

Odor: None — C7, C11
Spore Print: White
Gill Size: Narrow
Stalk Surface Above Ring: Smooth
Bruises: Yes

Implication

- We can translate a symbolic theory into a PDP network – productions as activities distributed across hidden units
- Perhaps PDP is not a "paradigm shift"
- Classical versus PDP debate requires more sophistication

Walter Schneider

What Kind Of Sophistication?

- Do other algorithmic equivalences exist?
- Do they map onto the same architecture?