


COGSCI 300  
**Week 5: Rules + Rosie**

Please turn off and put away all electronics.



1

## Wrongdoing Schema

1. Bad person
2. Bad intention
3. Bad act
4. Harm to victim

Examples: Cinderella, 3 Little Pigs, Little Red Riding Hood, Hansel & Gretel

2

## Wrongdoing Schema Explains


1. Knobe effect: people tend to think that actions that result in harm are intentional.
2. Moral dumbfounding (Haidt): no matter how acts such as incest are described, people tend to think of them as harmful.
3. Dyadic completion (Gray, JEP-G, 2015): wrong -> harm

3

## Objects of Moral Concern

Should we care about harm to fish, moles, puppies, robots?

What principle determines what we should care about?  
 Suffering, duties, emotional significance?



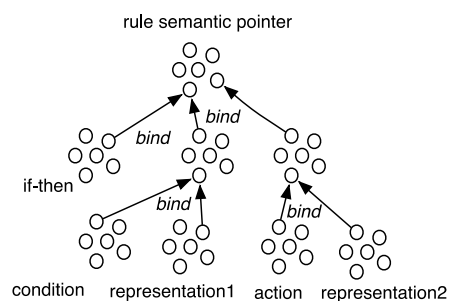
4

S. Murray, *J. Personality and Social Psychology*, 2015

1. IF conflict THEN relationship threat.
2. IF threats THEN relationship loss.
3. IF relationship loss THEN unhappy.
4. IF reduce threat THEN maintain relationship.
5. IF increase mutual dependence THEN reduce threat.
6. IF partner hurtful THEN be kind.
7. IF partner impedes goals, THEN resist devaluing partner.

5

Rules as Semantic Pointers



6

Rules as Semantic Pointers

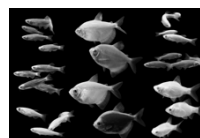
Advantages

1. Shows how complex inferences can be done by neurons.
2. Allows non-verbal representations in IF and THEN: sensory, motor, emotional.
3. Explains why rules can be unconscious.

7

Discussion Questions

1. Do fish suffer? How ethically significant is the answer?
2. What is the basis of moral rules?
3. How well do semantic pointers explain how rules are learned and used in problem solving, language, ethics, etc?



8

## What Rosie does



RObotic Soar Instructible Entity  
 Extension of SOAR, major cognitive architecture initiated by Allan Newell  
 Interacts with an external robotic environment  
 Learns to play games, e.g. TicTacToe  
 Learns definition of tasks  
 Learns by instruction

9

## How Does it Work?

### Representations:

Rules: if-then structures  
 Concepts: words for nouns, verbs, etc.  
 Scene graph (3-D) for spatial reasoning

### Procedures:

Interactive instruction  
 Learn legal actions, goals  
 Problem solving

10

## Discussion Question

How does Rosie compare with Watson, CYC, and Google cars?

Is Rosie a plausible model of learning by instruction?

11

## Rosie Strengths

1. Robotic connection to world
2. Learning by instruction, not just trial and error
3. Combines natural language understanding with problem solving
4. Interactive continuous learning
5. Learns many (15) different tasks

Note: only humans teach!

12

## Rosie Limitations

1. Limited syntax and communication
2. Does not scale to larger games
3. Limited problem solving
4. Limited in recursive binding, imagery, analogy,, emotions, consciousness, creativity
5. Could Rosie teach Rosie?

13

## Rosie vs. Humans

1. Advantages of Rosie: no fatigue, boredom, distraction
2. Advantages of humans: multimodal concepts, emotions for motivation, recursive binding, creativity, context

14