Israel and Behavioural Economics

Adventures in Decision Making at the Asia Pacific Meeting of the Economic Science Association and the Centre for the Study of Rationality Workshop on Regret, Emotions, and Decision Making

University of Waterloo

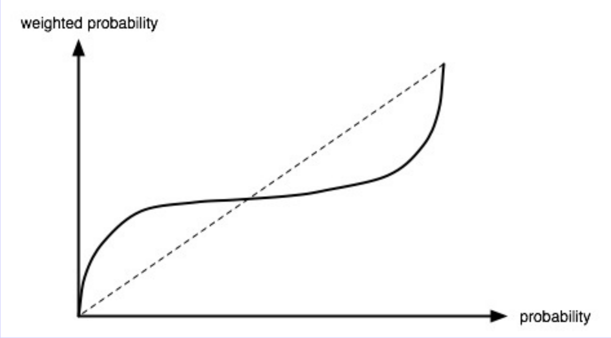
March 23-29, 2009

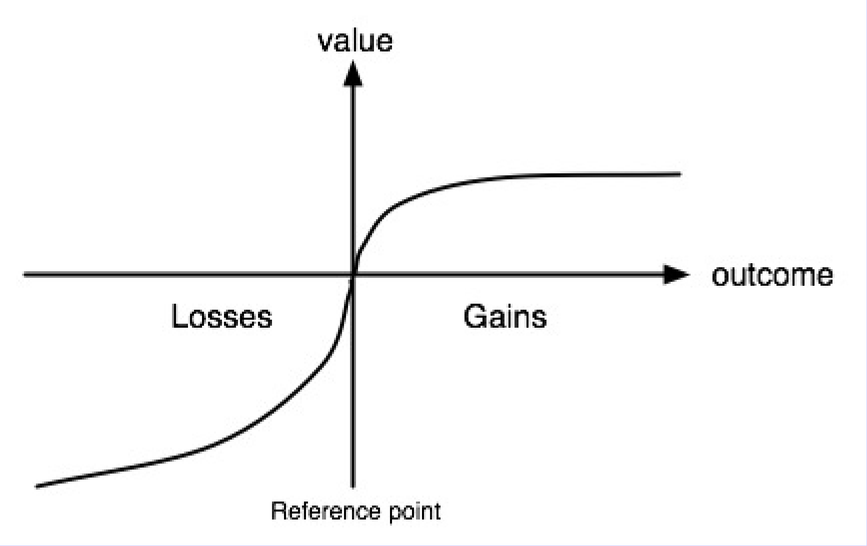
Behavioural Economics

- What are these people interested in?
  - Decision making, Game theory, etc.
  - Prisoner's Dilemma
  - Ultimatum Game
  - Binary Choice between Gambles
  - Auctions, Markets, Common Goods, Trust, etc.
- Common tools?
  - Questionnaires, Lab experiments with real money
  - Prospect Theory

Prospect Theory

- Kahneman & Tversky (Nobel Prize in Economics)
  - Expected Utility Theory doesn't work $$U = \sum p_i x_i$$
  - Need to adjust values and probabilities
  - v(x) requires a reference point and makes losses worth more than gains (loss aversion) $$U = \sum w(p_i) v(x_i)$$
  - w(p) handles overreacting to small probability events





Prospect Theory: Evidence

- Loss Aversion
  - Choose A: +200 or B: 67% 0; 33% +600 [A>B]
  - Choose A: -400 or B: 67% -600; 33% 0 [A<B]
- Overreact to small probabilities
  - Buying lottery tickets
  - Buying insurance
  - Cumulative Prospect Theory changes w(p) to be w(p,x) where it overweights events that are both rare and extreme

Some General Notes

- Emphasis on verbal descriptions of options
- Descriptive models
  - Measure behaviour, find curves that fit individuals
  - Finding patterns (loss aversion), use for explanation
- But
  - No mechanism
  - Post-hoc fitting, not prediction in new domains

Prospects in other domains

- Does this apply to things other than money?
  - Yes
  - Life expectancy, weight, waiting times, pain
- Does this apply to small versus large values?
  - Yes
  - Pennies vs thousands of dollars (hypothetically)

Relative Thinking

- $125/$25 jacket/calculator question
  - Will drive 20min to save $5 on $25, but not $5 on $125
- New domain: boring questions
  - Initial $5, then $0.15 per really really boring question
  - Compare to initial $15
  - Predictions
    - Prospect theory: no diff
    - Reciprocity: more for $15
    - Relative: more for $5
  - “found a slight tendency for relative, but not statistically significant”

(Note: Gigerenzer, 1998)

Emotions: Fear vs anger

- How do emotions affect decisions?
  - Most work on emotions in this area just looks at +ve vs -ve emotions
  - Fear vs anger? (abstract -ve vs concrete -ve)
  - Risk aversion: fear>happiness>anger
- Does the type of gamble matter? Yes
  - lotteries give fear>happiness>anger
  - Interactions give fear<happiness<anger
  - Note that for both cases -ve=+ve

Reconsidering Loss Aversion

- Ido Erev
  - “The assertion that losses loom larger than gains is one of the best known implications of prospect theory.”
    - I'm not sure it's an implication – feels more like an assumption
  - Initial data
    - 0 vs 50% -1000; 50% +1000
    - +1000 vs 50% 0; 50% +2000: didn't become more attractive either as a hypothetical or as real money
  - What's happening?

Reconsidering Loss Aversion

- Framing effect:
  - If describe gamble, ask “would you do it”, makes status quo more attractive
  - If describe both options, pick one, gamble becomes more likely
- Mental exhaustion effect:
  - Loss aversion found in long experiments (>50 choices)
  - Re-analyzed raw data, found no loss aversion at beginning (<10 choices)
  - Loss aversion also goes away with feedback
    - Comes back with large magnitudes

Prediction Competition

- Why was I there?
  - Ido Erev organized a competition to see whether choice models could predict
  - 3 Conditions
    - Describe the gambles, then pick
    - Explore the gambles for a while, then pick
    - No description, just keep picking 100 times
  - I won the last condition (most predictive model)

Repeated Binary Choice

Repeated Binary Choice

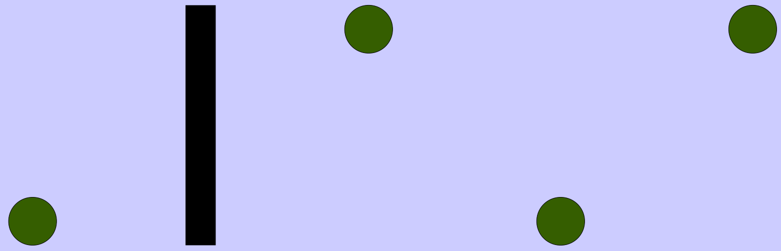
8	

Repeated Binary Choice

- Safe vs. Risky Options
  - One button always gave a reward M
  - Other gave reward H with probability p, otherwise L
- 120 Conditions
  - Varying L, M, H, and p
  - Values between -30 and +30, rare events (p< .1 or p>.9) in about 2/3 of the problems
- Generalization
  - Only given data from 60 Conditions (proportion of risky choices over 100 trials, 20 subjects)
  - Need to predict results from other 60 conditions.

Competition Results

- Description: won by prospect theory
- Explore, then pick: won by combining models
  - If have multiple models and no theoretical reason to choose one over the other, then average them all
  - No weighting



  - But there's no way that's what people are doing

Cognitive Modelling

- Cognitive Science:
  - How does the brain work?
  - Both what it does and how it does it
  - Identifying the underlying mental mechanisms
- Why would this be useful?
  - If understand components and how they fit together, should be able to make new models for new situations
  - It should be possible to make a predictive model without running any experiments

Mechanistic Models

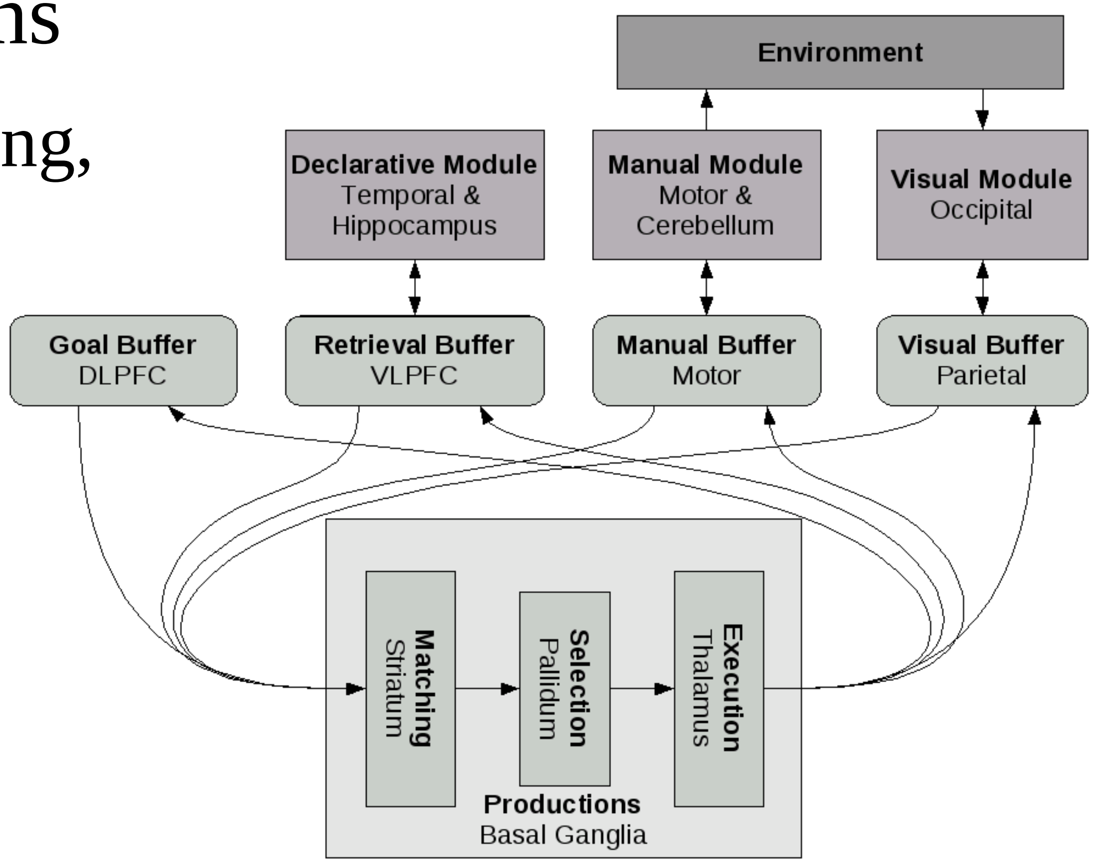
- Constraints on cognitive models
  - Process models (consistent flow of time)
  - Components correspond to brain components
- Benefit of constraints
  - Claim: doing this leads to better predictions
  - Predictions across many domains (reaction time, fMRI, lesions, interventions, learning trajectory, etc.)

Cognitive Architectures

- Avoiding Reinventing the Wheel
  - Cognitive models are highly complex
  - A lot of parameters, possibilities
  - Solution: use common components, parameter values
- Benefit of constraints
  - Acts somewhat like a Bayesian prior in new tasks
  - Biases models towards components and algorithms that have previously shown to be successful
  - Reasonable to believe that they will be useful here too

ACT-R

- The most widely used cognitive architecture
  - 15-30 years old; about 100 active researchers
- Used in many domains
  - mathematical reasoning, serial recall, human computer interaction, semantic priming, n-back, sleep deprivation, driving while dialing, etc., etc.



Cognitive Architectures

- Why use ACT-R?
  - Usually used in situations with multiple types of data
    - Accuracy, reaction time, fMRI, task interference, experience
- Validated Components
  - We have reason to believe the brain can use these components in other tasks
  - Reasonable to believe it could also use them in this task
  - Acts somewhat like a Bayesian prior

The Model

- Treat repeated choice as a memory task
  - Recall what happened after pressing each button
  - Pick the biggest
- Constrain recall based on context
  - If I just pressed AA, only consider previous experiences with AAA and AAB
  - History of 2 from rock-paper-scissors, baseball

The Model

- ACT-R Memory
  - Odds of a memory being needed decay as a power law over time (Anderson & Schooler, 1991)
  - Each memory has activation *A*
  - Let t_i be the times since this memory was seen

$$A = \ln \sum t_i^{-d} + \epsilon(s)$$

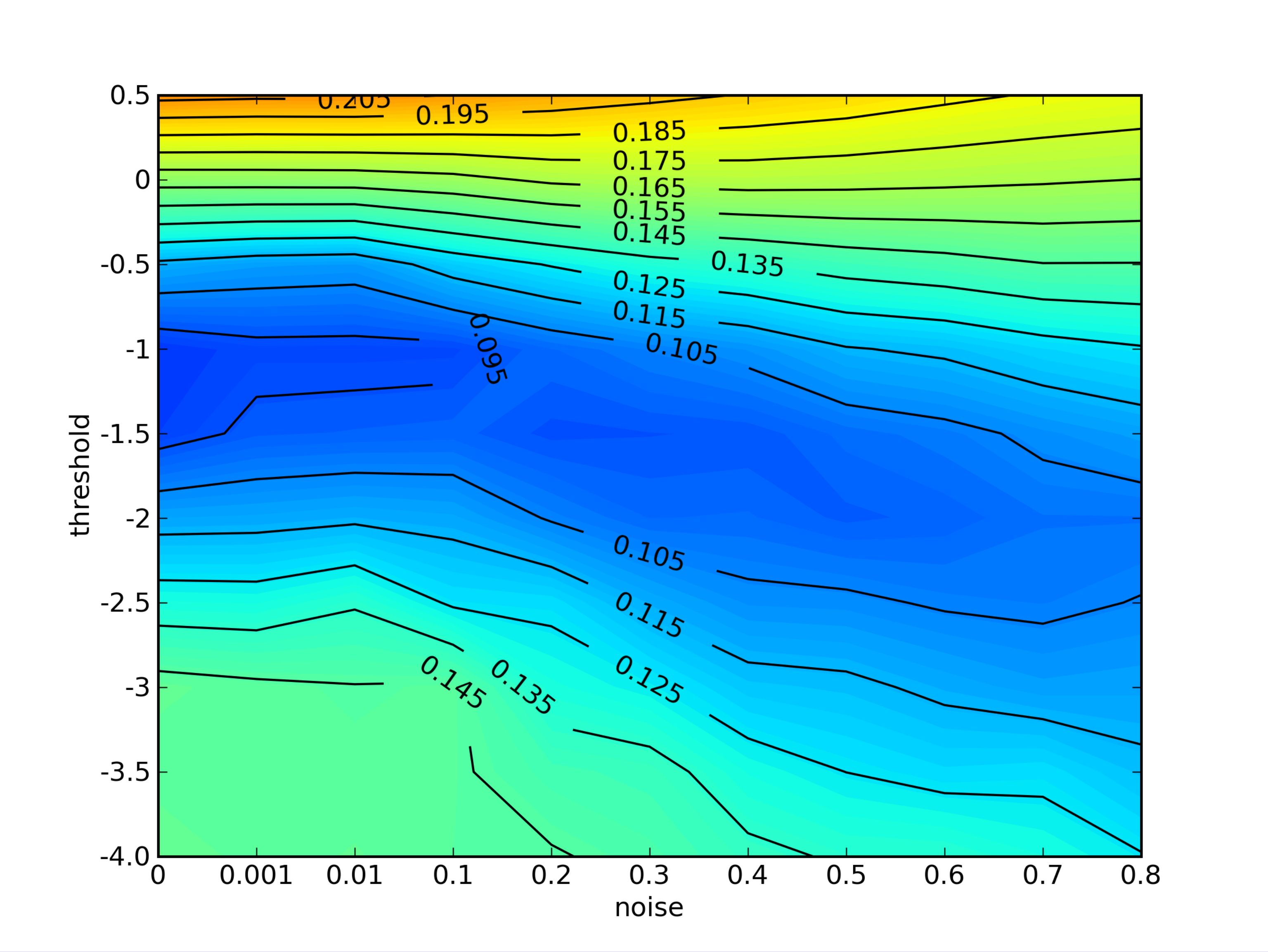
  - Recalls the item with highest *A*, if above threshold T
  - Magnitudes are not well distinguished
    - Blend together, linearly weighted by A
  - Only one recall at a time: $time = Fe^{-A}$

Model Parameters

- Decay rate *d*
  - Always 0.5
- Size of context
  - 2 is consistently best
- Noise *s*
  - Usually between 0.2 and 0.5
- Threshold *T*
  - Almost always between -3 and 0
- Two parameter model

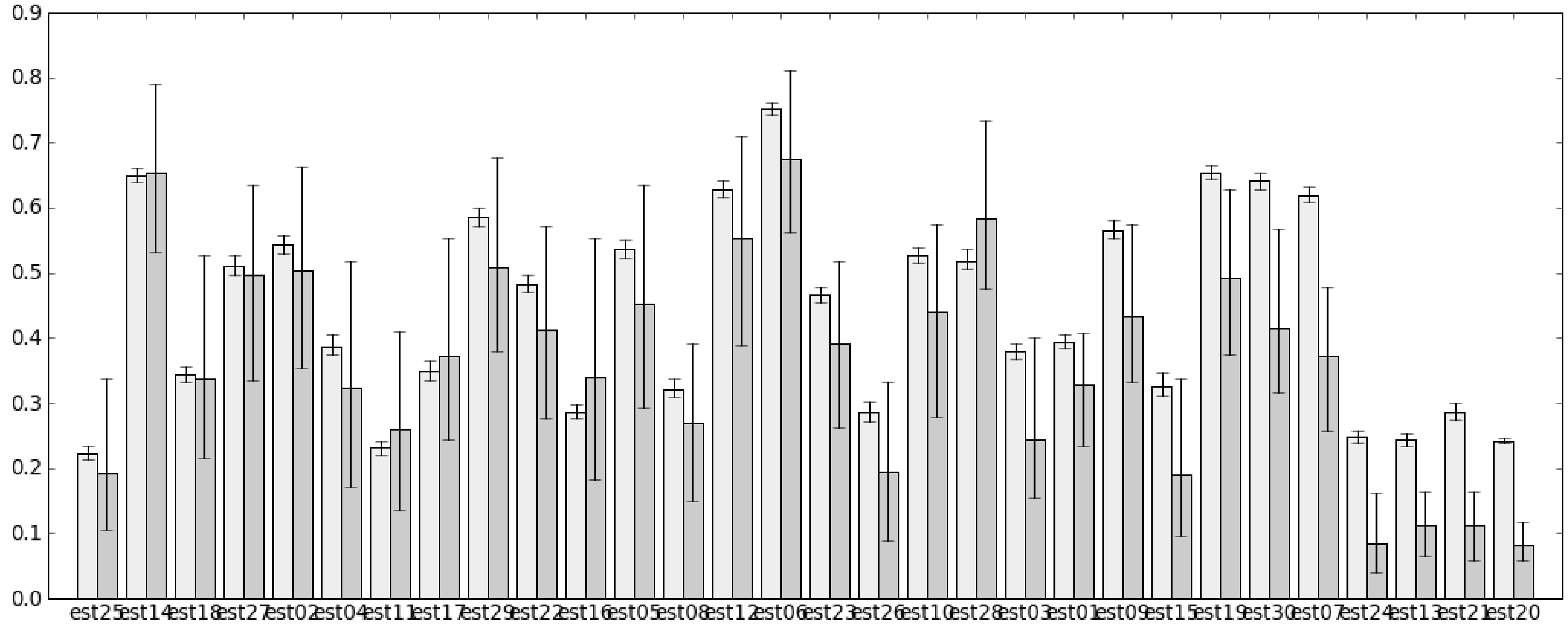
Evaluation

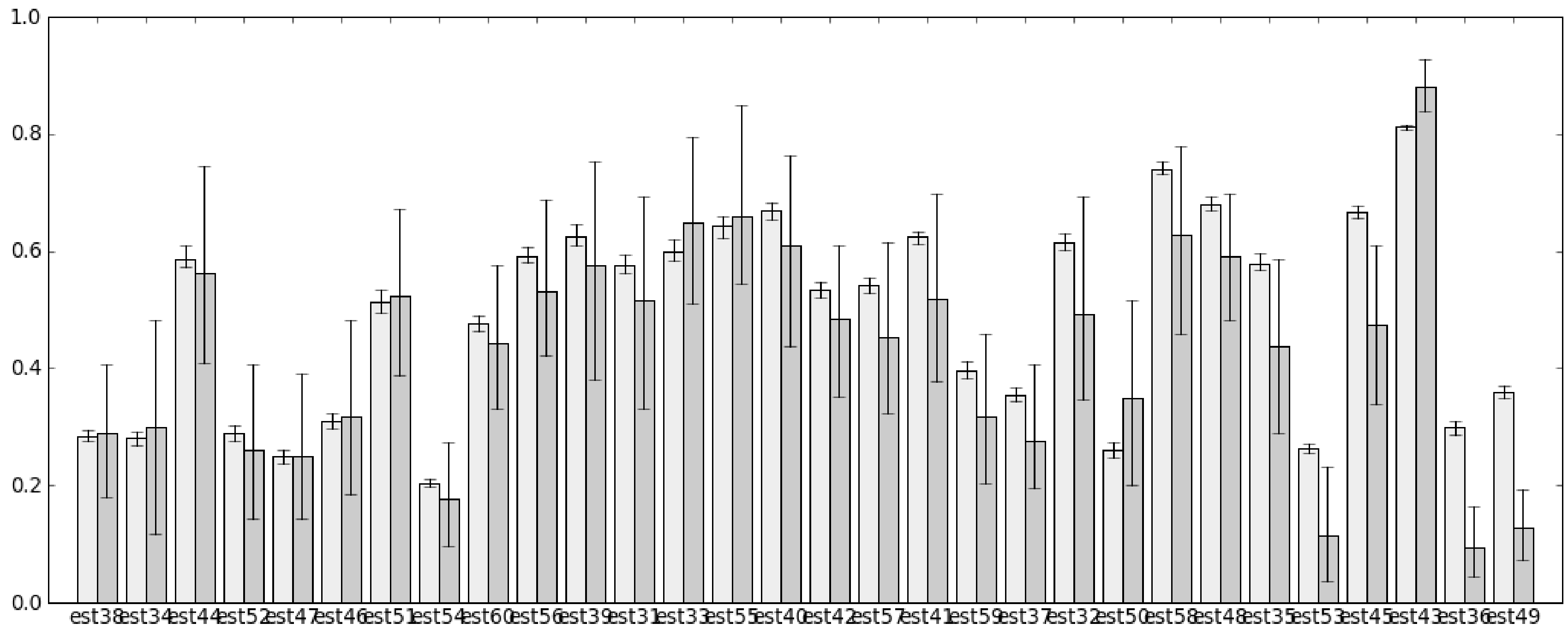
- RMSE to human data on 60 conditions



Evaluation

- Best performance is s=0, T=-1
  - This is not a common value for s
  - What is happening here?
- Let's look at model and human performance across the 60 conditions

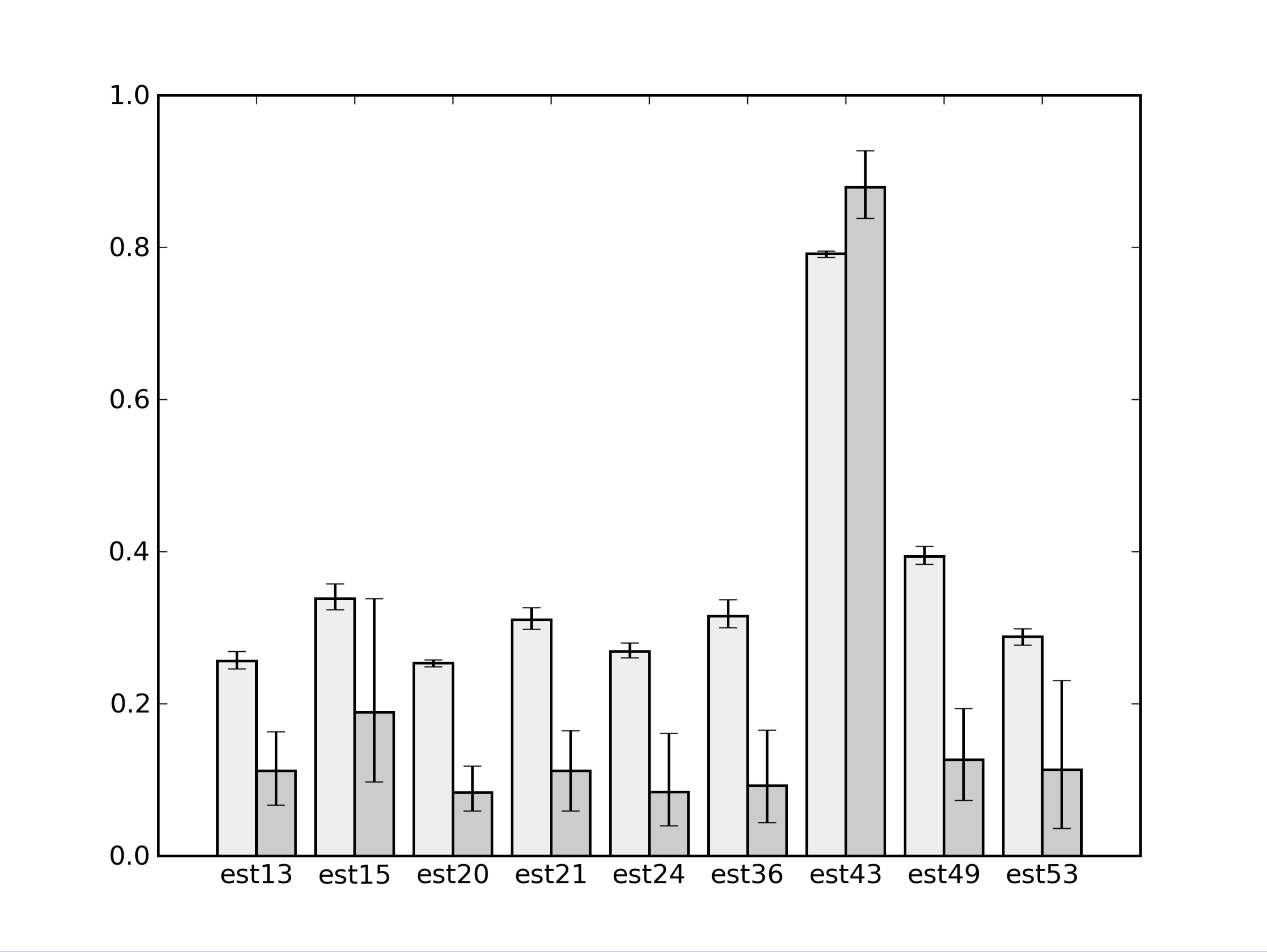

0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
est25 est14 est18 est27 est02 est04 est11 est17 est29 est22 est16 est05 est08 est12 est06 est23 est26 est10 est28 est03 est01 est09 est15 est19 est30 est07 est24 est13 est21 est20


0.0
0.2
0.4
0.6
0.8
1.0
est38 est34 est44 est52 est47 est46 est51 est54 est60 est56 est39 est31 est33 est55 est40 est42 est57 est41 est59 est37 est32 est50 est58 est48 est35 est53 est45 est43 est36 est49

Evaluation

- Some measures consistently don't match
  - Can't get them to match while other ones do too
  - On the other measures, there is no statistically significant difference between model and empirical data
- Conclusion
  - The model does not accurately capture a few conditions
  - Something else may be happening here
  - Should not try to average over those models
    - Separate groups of conditions

Extreme Conditions

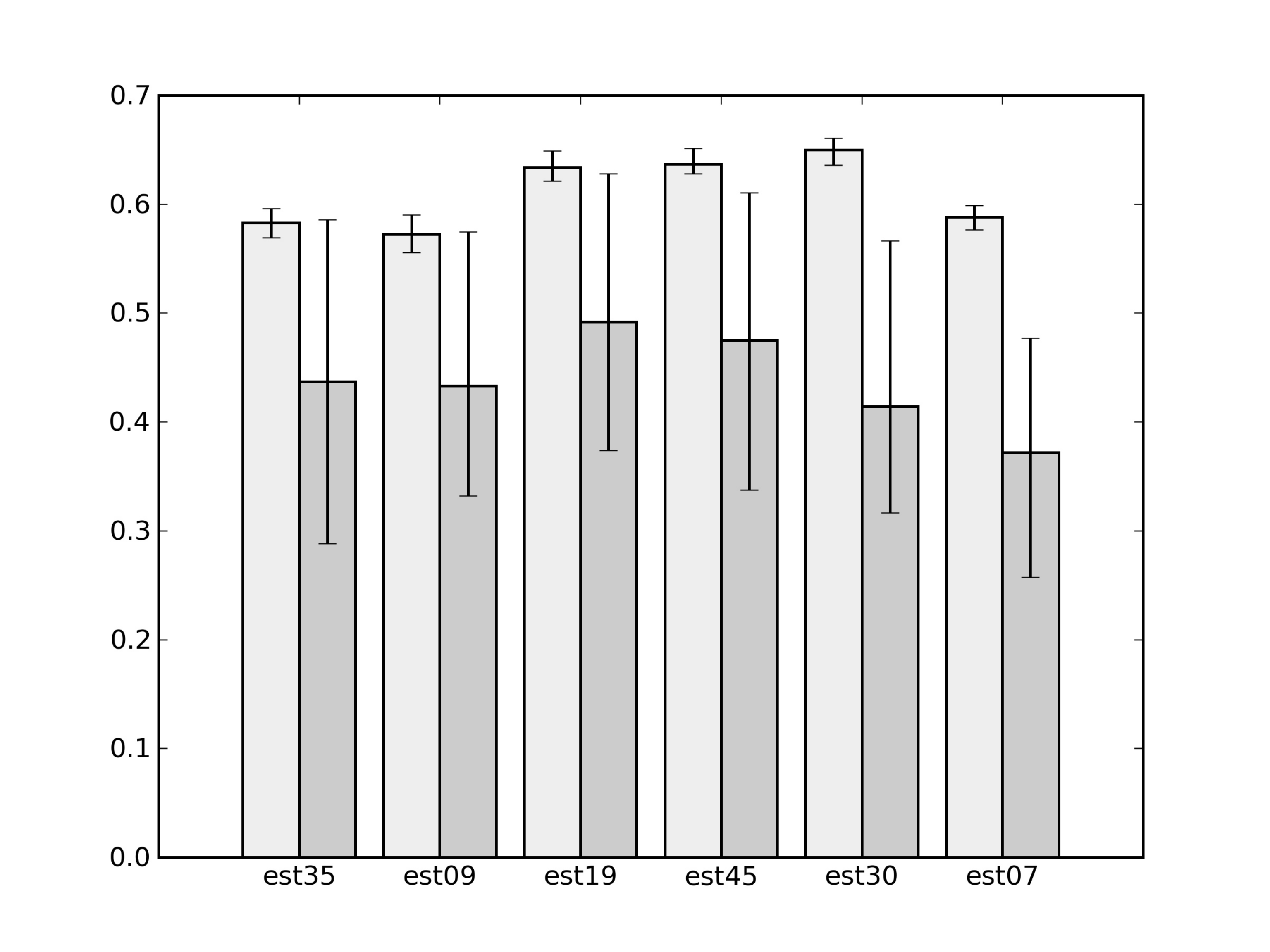


Extreme Conditions

- Conditions
  - All of the situations where people exhibited extreme behaviour
  - Perhaps there is a strategy shift when people notice one button is consistently better?

#	H	p	L	M
13	-2	0.05	-10.4	-9.4
15	-8.9	0.08	-26.3	-25.4
20	-4.3	0.6	-16.1	-4.5
21	2	0.1	-5.7	-4.6
24	9.2	0.05	-9.5	-7.5
36	5	0.08	-9.1	-7.9
43	22.9	0.06	9.6	9.2
49	13.4	0.5	3.8	9.9
53	25.7	0.1	8.1	11.5

60/40 Error Conditions



60/40 Error Conditions

- Not sure what is happening here
  - Not even sure if it's a real effect
  - 95% confidence intervals will be wrong 5% of the time
  - Something to do with rare negative outcomes
    - Model likes risk in these conditions more than people do

#	H	p	L	M
35	3	0.93	-7.2	2.2
9	-5.7	0.95	-16.3	-6.1
19	-6.5	0.9	-17.5	-8.4
45	2.8	0.8	1	2.2
30	3	0.91	-7.7	1.4
7	-5.6	0.7	-20.2	-11.7

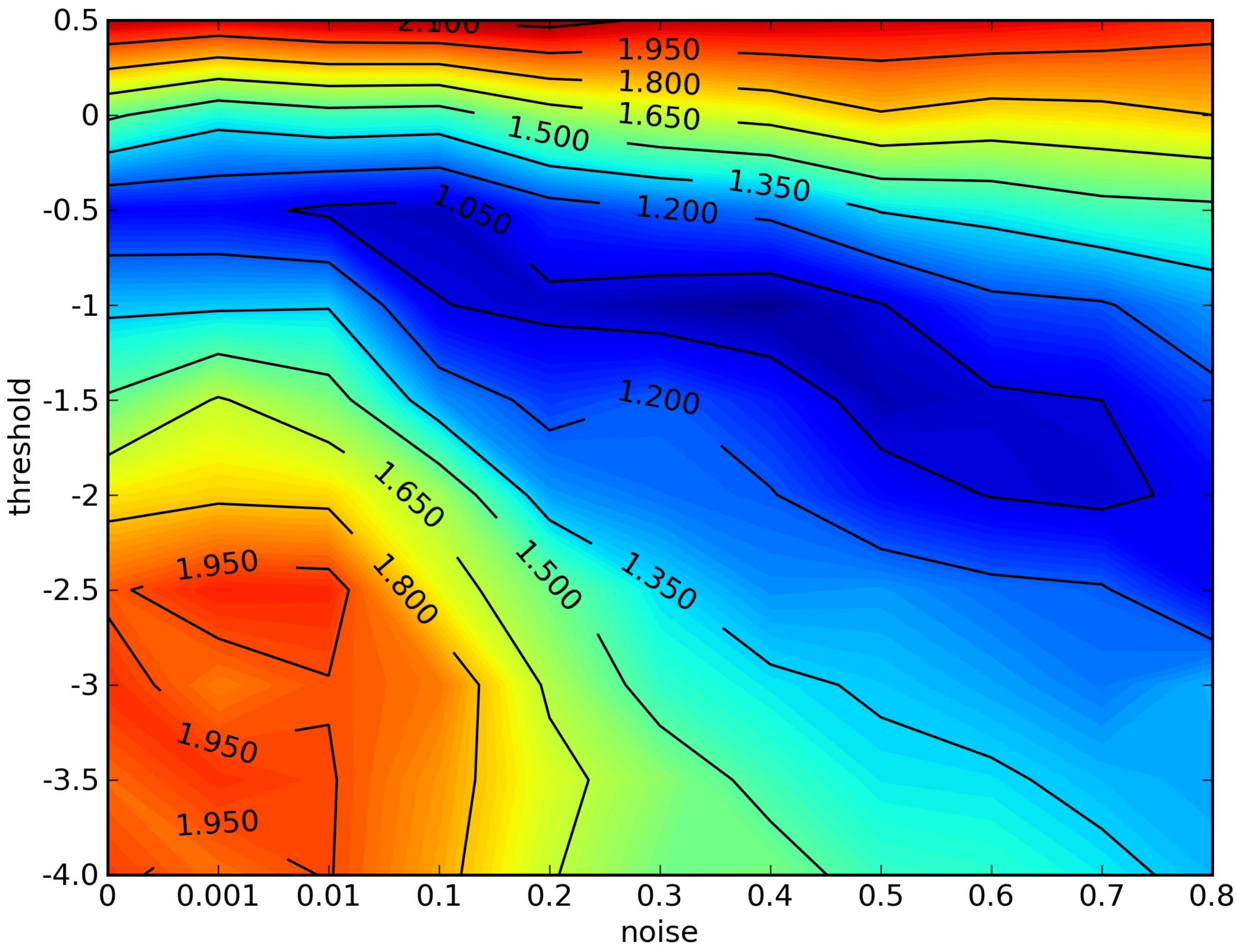
Conservative Modelling

- The model is not perfect
  - Instead of averaging over all the conditions, identify the conditions for which the model is good
- What do we mean by a good match?
  - Let's continue looking at confidence intervals
  - After removing the consistently bad conditions, what is the next worst condition?
  - Measure this for each parameter setting

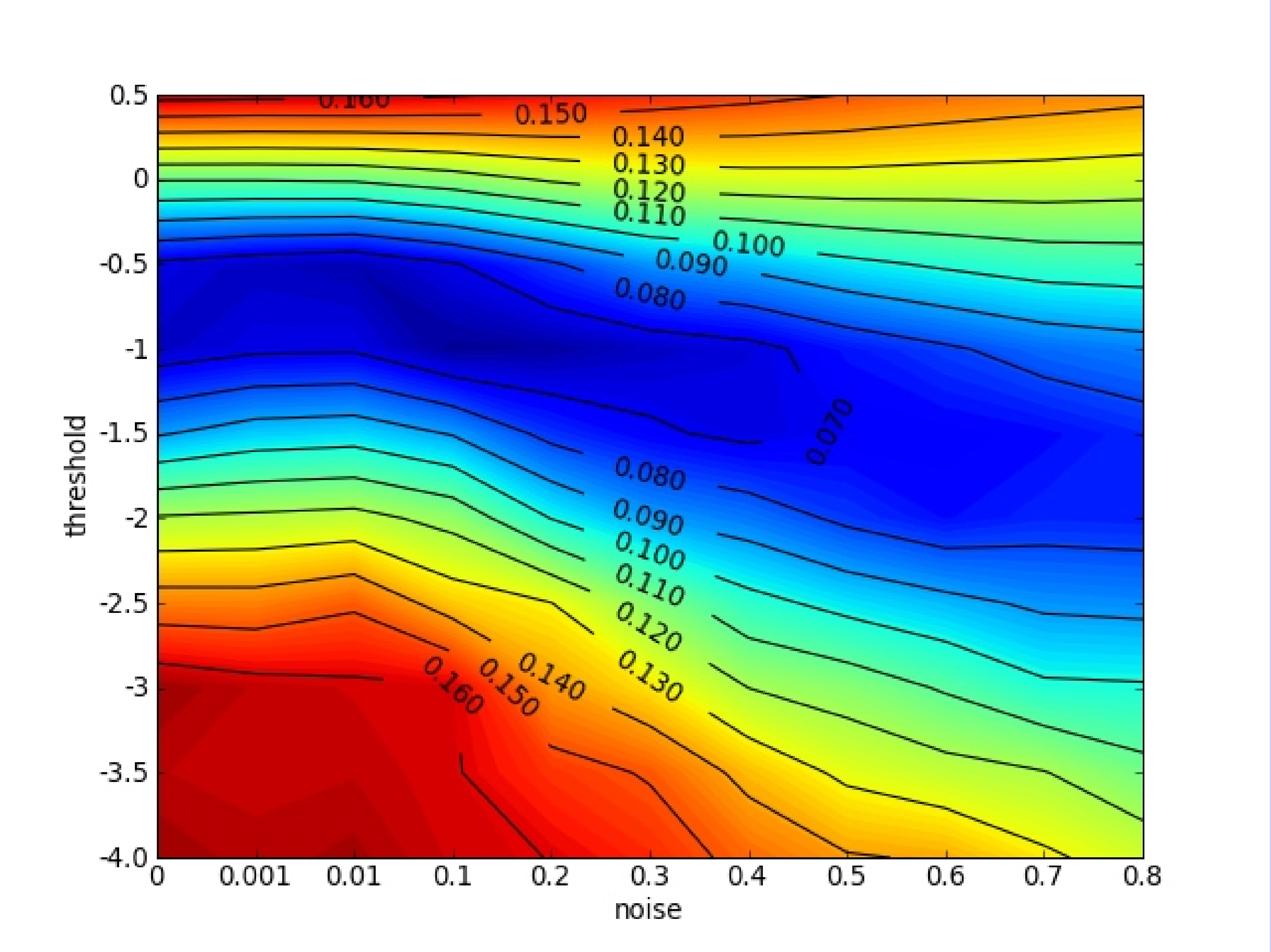
Scaled Maximum Likely Difference

- Likely Difference
  - How bad could the model be, but still be consistent with the observed confidence intervals
- Scaled Likely Difference
  - Linearly scaled so that <1 means no statistically significant difference between model and data
- Scaled Maximum Likely Difference
  - Maximum of SLD values across all conditions
  - Highly conservative worst-case assessment of model quality

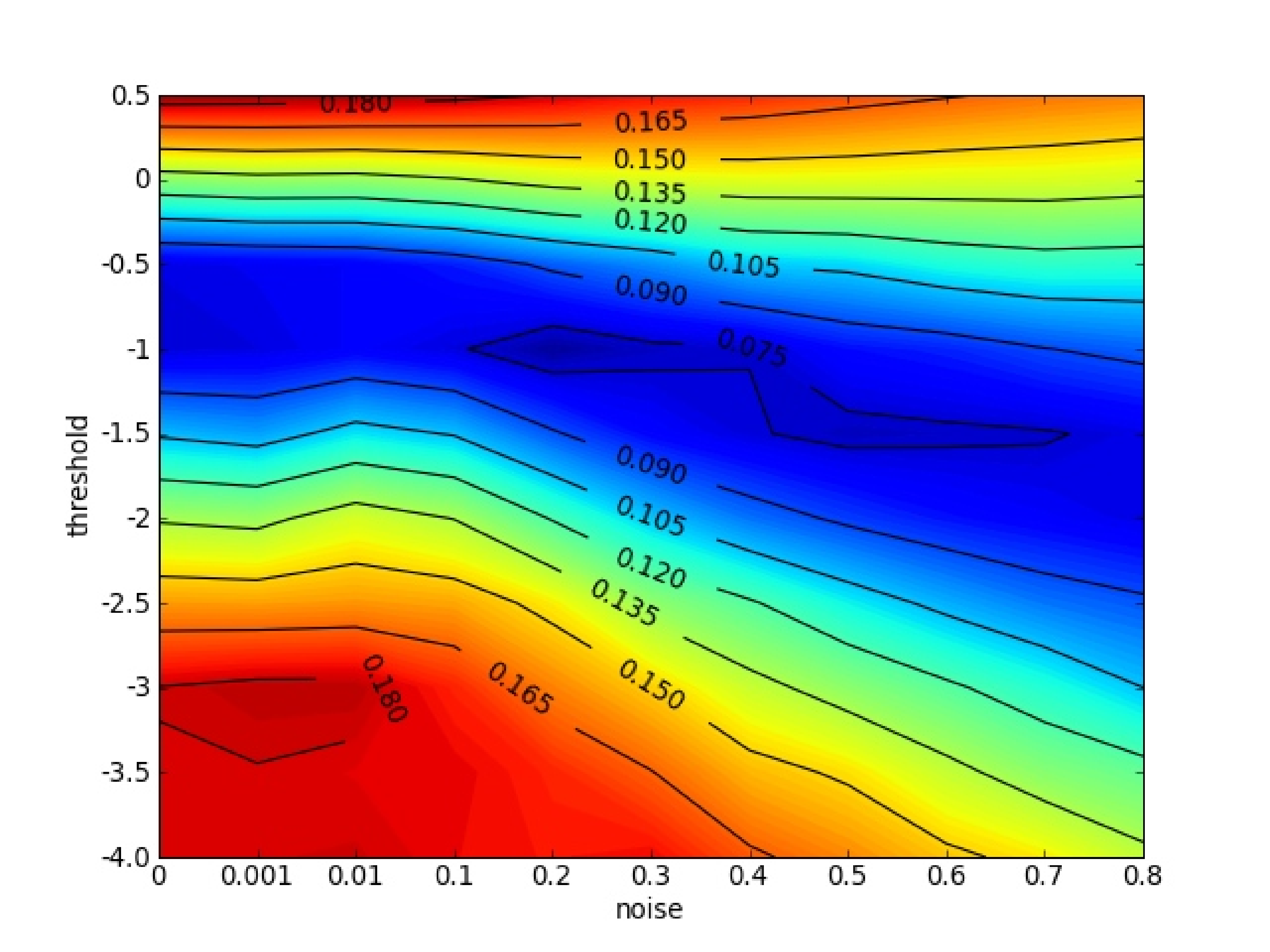
Scaled Maximum Likely Difference



RMSE with conditions removed



Testing Data



Conclusions

- Useful to base a model on existing cognitive theory
  - Constrains model parameters and functions to ones that have proven to be good in the past
  - Acts as a Bayesian prior to give you more predictive accuracy than a small set of task-specific empirical data
- Useful to identify conditions the model does not handle well
  - No averaging over good and bad conditions
  - Would not have won otherwise
  - Helps direct future research

Heuristics and Metacognition

> Common sense says, we lose our fortune, are sorry and weep; we meet a bear, are frightened and run; we are insulted by a rival, and angry and strike. The hypothesis here to be defended says that this order of sequence is incorrect and that the more rational statement is that we feel sorry because we cry, angry because we strike, afraid because we tremble. Without the bodily states following on the perception, the latter would be purely cognitive in form, pale, colourless, destitute of emotional warmth. We might then see the bear, and judge it best to run, receive the insult and deem it right to strike, but we could not actually^/ afraid or angry, (p. 190)

- Asher Koriat
  - William James (1884):
  - Why the feeling? To help guide behaviour
    - Emotions have a monitoring->control role
    - e.g. Koriat & Goldsmith, 1996: confidence predicts likelihood of answering a question, even if confidence is uncorrelated to accuracy
- Apply to memorizing a list of items
  - Will choose to spend more time on the harder ones
    - How do you know which ones are harder?
    - By the amount of time spent on them

Heuristics and Metacognition

- Very circular
  - So let's play with their time perception
    - Very visible clock with slight speed changes
    - If clock runs slower, less confident in performance
- Good heuristic
  - Effort is data-driven, so time is a good diagnostic
  - Metacognitive processes are parasitic on underlying cognitive ability
    - Kids only start using the cue around grade 3

Heuristics and Metacognition

- But there's a cycle
  - control->monitor->control->monitor->...
  - When studying, you can choose to allocate more time to some items
  - Have some items worth 1pt, others worth 3pts
    - Across: +ve correlation (judgment vs accuracy)
    - Within: -ve correlation (unless there's lots of time pressure, which starts making it goal-driven within the 3pt items)

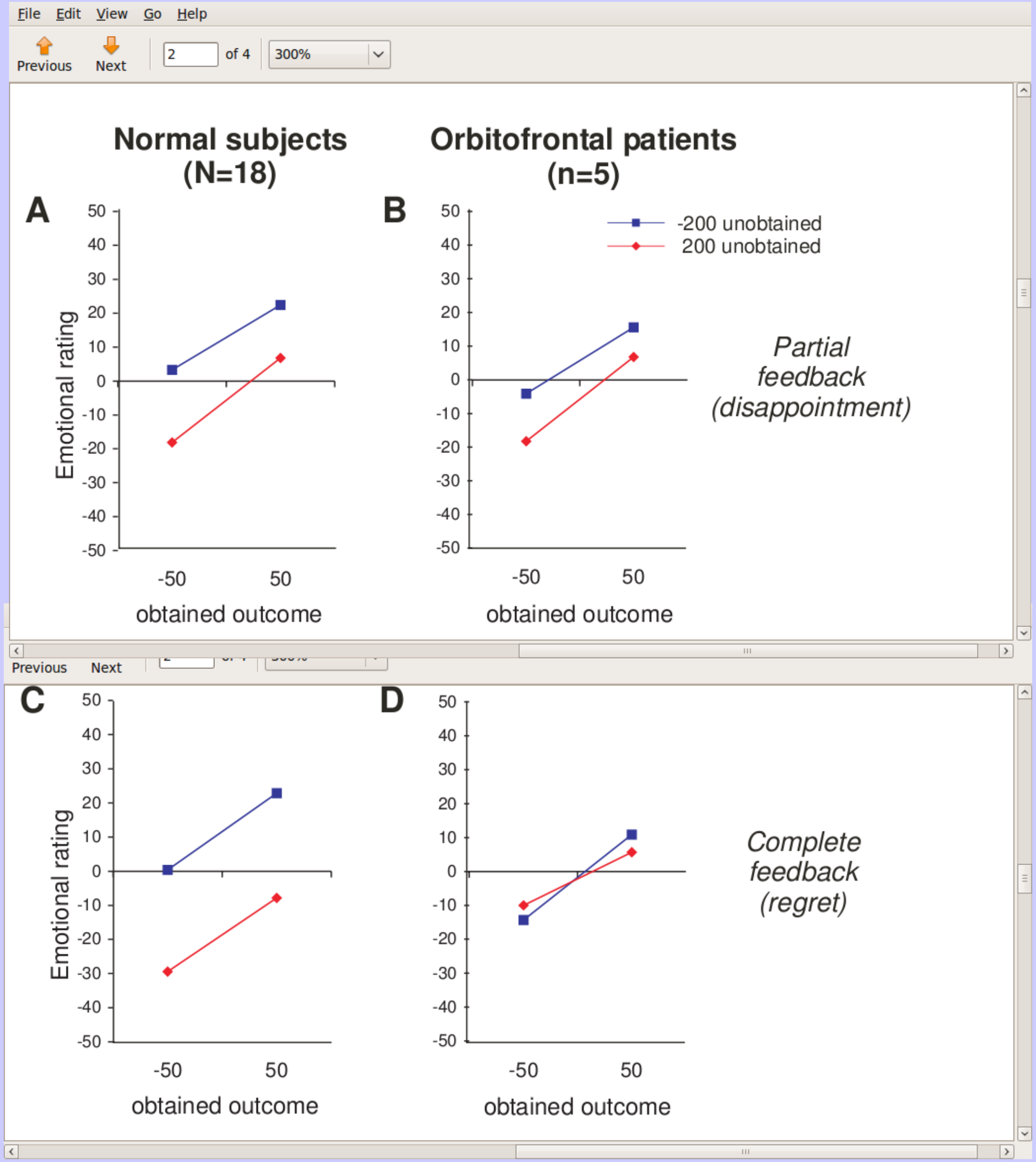
Regret: Generating Counterfactuals

- Marcel Zeelenberg
  - Given a scenario where something went badly:
    - Generate 0/2/8 counterfactuals how it could be better
  - Ask how much regret you would feel
    - 0=2>8
  - If told that 2 is easy and 8 is hard
    - 0=2=8
  - Experienced ease of generating counterfactuals is used to estimate regret
  - But it won't be used if told it's not a good indicator

Regret: Fictive learning

- Giorgio Coricelli
  - Orbitofrontal cortex: representation of relative reward
  - Can adjust by having feedback of what you would have gotten with the other choice
  - Also look at patients with lesions to this area

Regret: Fictive learning



Regret: Fictive Learning

- Orbitofrontal patients can't do the counterfactual
  - So the don't feel regret, and so don't use it to adjust behaviour
- Two learning mechanisms
  - Expected reward-based (prediction error)
    - Obtained-expected (expected=Ap+B(1-p))
  - Fictive learning
    - Obtained-unchosen (A-B ignoring probability)
  - Regression analysis of decisions
    - Normals use expected and regret
    - Orbitofrontals just use expected