Inverse Preference Elicitation

University of Michigan Addiction Research Center - 17 February 2011

Dan Lizotte
Postdoctoral Fellow
Department of Statistics

With Michael Bowling, Susan Murphy
University of Alberta, University of Michigan

UNIVERSITY OF MICHIGAN

Outline

- Part I
  - Motivation: Symptoms and Side-Effects in Schizophrenia
  - Background: Predictive Models and Optimal Decision Rules
  - Contribution: Inverse Preference Elicitation
- Part II
  - IPE for Sequences of Actions
  - Results: Exploratory Analysis of the CATIE Antipsychotic trial
  - Discussion and Future Work:
    - Experimental evaluation using Mechanical Turk
    - Other extensions

Motivation:
Symptoms and Side-Effects in Schizophrenia

Motivation:
Symptoms and Side-Effects in Schizophrenia

- Many treatments available for treating schizophrenia (dozens)
- Evidence-based medicine would look at predicted outcomes, recommend a **treatment**
- *At least* two important objectives:
    - Maximize symptom reduction, minimize weight gain

Motivation:
Symptoms and Side-Effects in Schizophrenia

- Many treatments available for treating schizophrenia (dozens)
- Evidence-based medicine would look at predicted outcomes, recommend a **treatment**
- *At least* two important objectives:
    - Maximize symptom reduction, minimize weight gain
- Treatments that provide the best symptom reduction induce the worst weight gain, and vice-versa

Motivation: Symptoms and Side-Effects in Schizophrenia

- Many treatments available for treating schizophrenia (dozens)
- Evidence-based medicine would look at predicted outcomes, recommend a **treatment**
- *At least* two important objectives:
    - Maximize symptom reduction, minimize weight gain
- Treatments that provide the best symptom reduction induce the worst weight gain, and vice-versa
- Different doctors and patients have very different preferences about relative importance of outcomes

Motivation:
Symptoms and Side-Effects in Schizophrenia

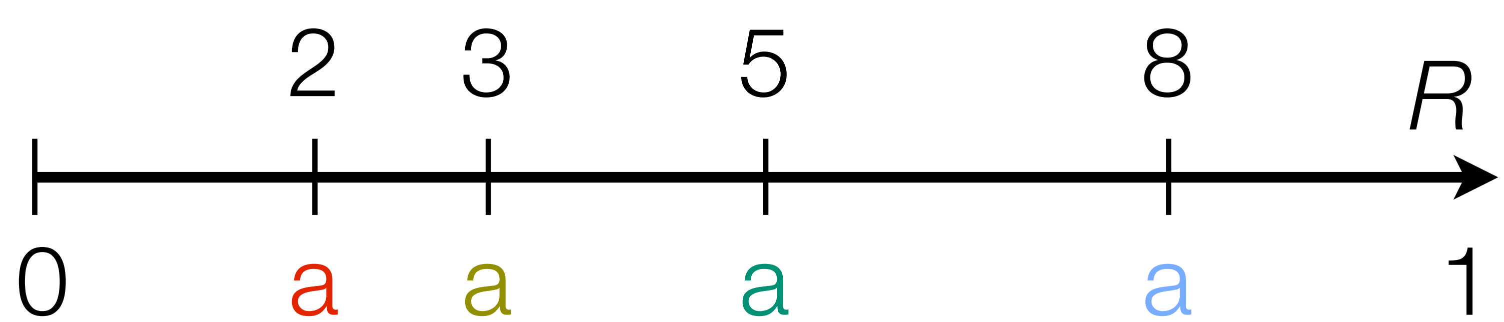
- Many treatments available for treating schizophrenia (dozens)
- Evidence-based medicine would look at predicted outcomes, recommend a **treatment**
- *At least* two important objectives:
  - Maximize symptom reduction, minimize weight gain
- Treatments that provide the best symptom reduction induce the worst weight gain, and vice-versa
- Different doctors and patients have very different preferences about relative importance of outcomes
  - **How can we recommend a treatment that accommodates these preferences?**

Outcome Predictions and Decision Rules - Single Outcome

- Identify an outcome of interest (**reward**) R,
  predictive patient features (**state**) S,
  and a set of treatments (**actions**) A
- Construct a predictive model
  - Input: (S, A) Output: Prediction of R
- Could be done by regressing R on (S, A) for example
- e.g., have (S, A, R) for each individual, A is randomized

Optimal Decision Rule - Single Outcome

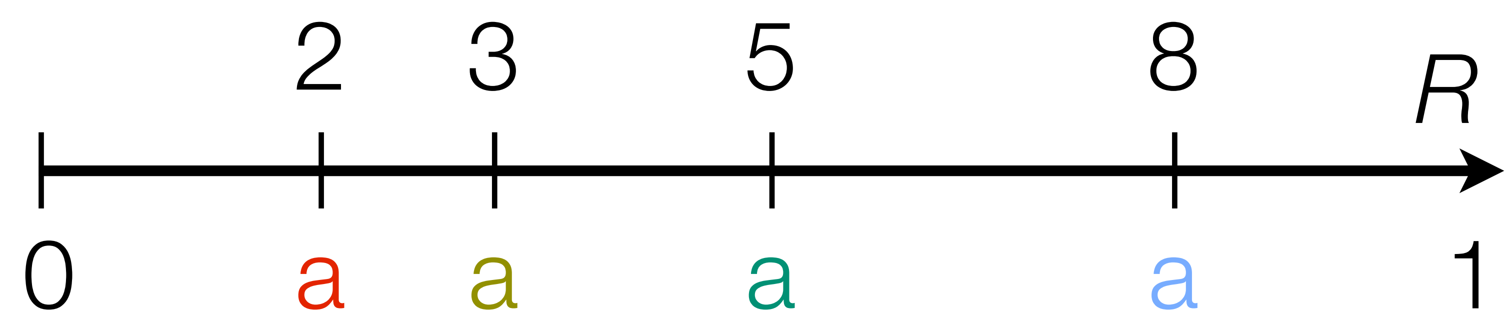
Predicted R for patient with S=s



Optimal Decision Rule - Single Outcome

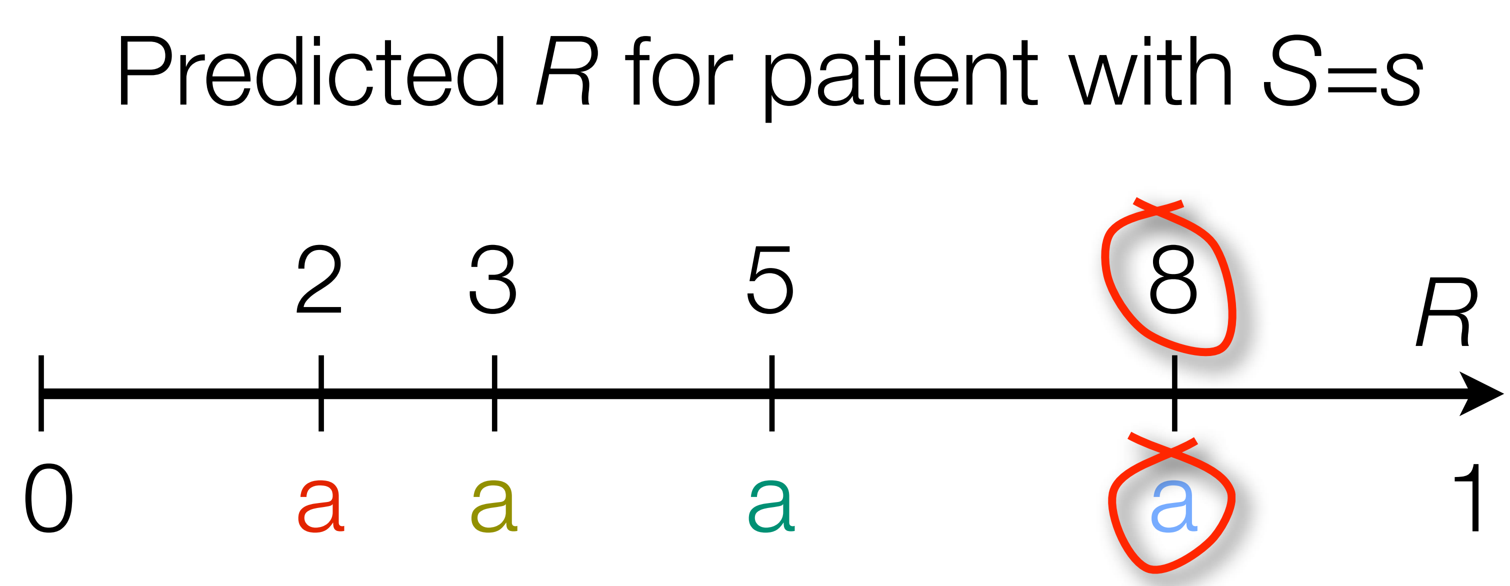
- A model makes predictions of the reward of a patient with state *s* under 4 different actions, a, a, a, and a.

Predicted *R* for patient with *S*=*s*



Optimal Decision Rule - Single Outcome

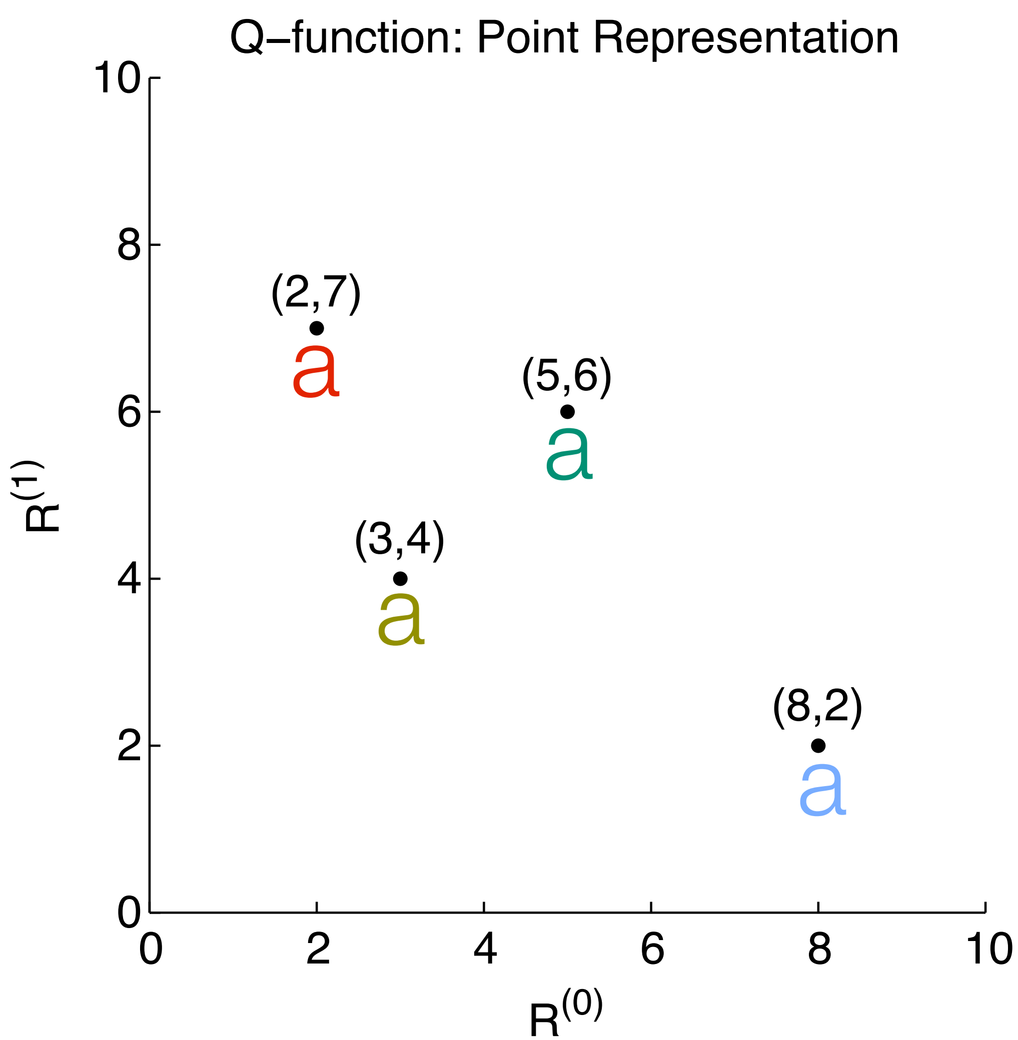
- A model makes predictions of the reward of a patient with state *s* under 4 different actions, a, a, a, and a.
- For the patient with the predictions shown below, action a looks best, with a predicted reward of 8. (Higher rewards are better.)



Optimal Decision Rule?

- **Two models** make predictions of **two different rewards** of a patient with state *s* under 4 different actions, a, a, a, and a.

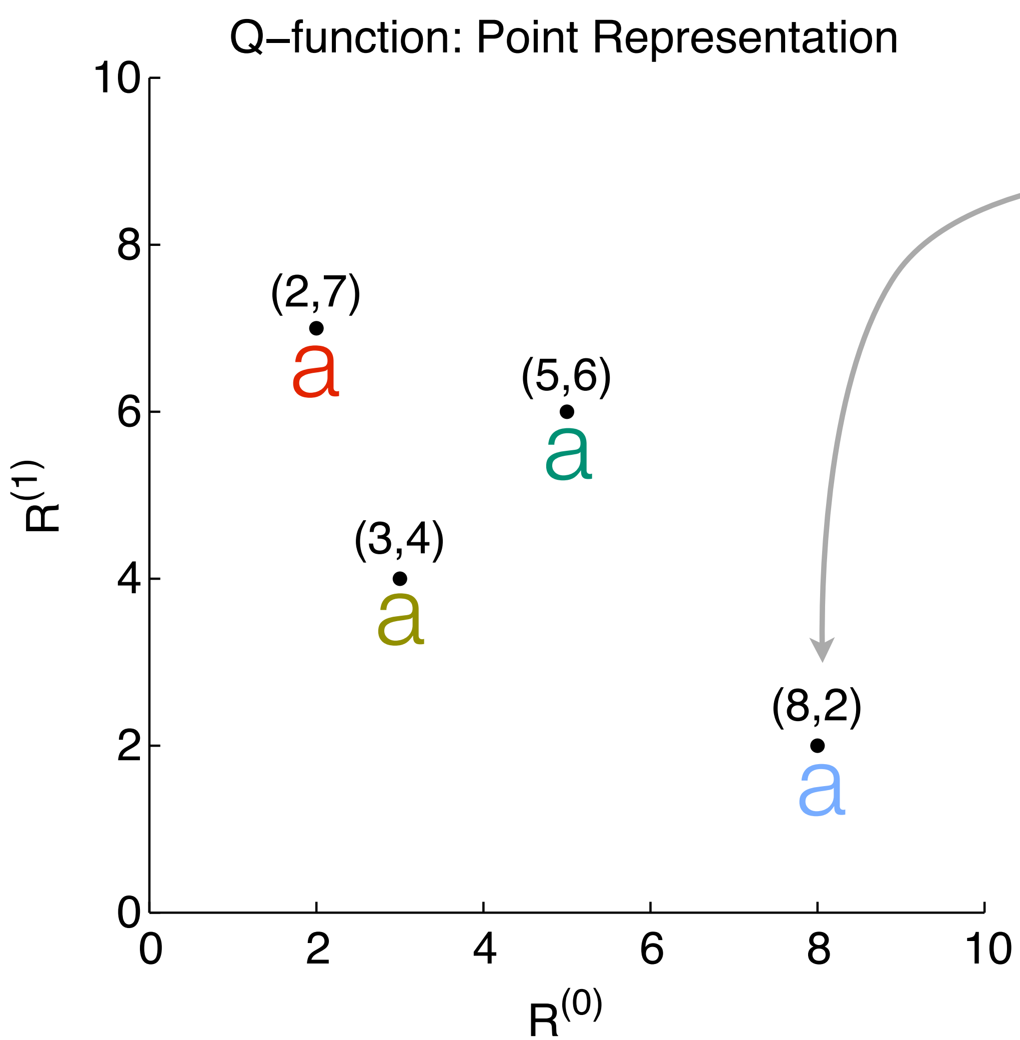
Predicted ($R^{(0)}$, $R^{(1)}$) for patient with $S = s$



Optimal Decision Rule?

- **Two models** make predictions of **two different rewards** of a patient with state *s* under 4 different actions, a, a, a, and a.

Predicted $(R^{(0)}, R^{(1)})$ for patient with $S = s$

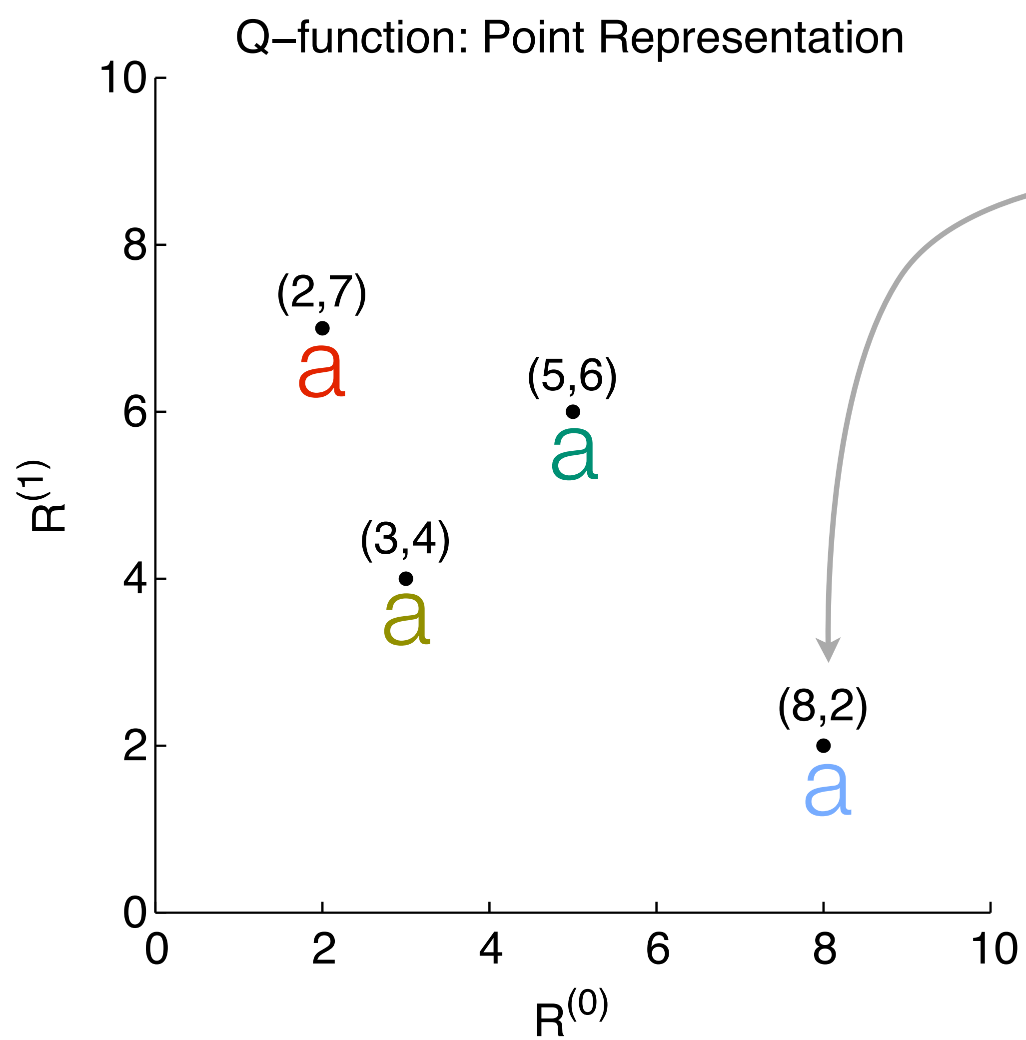


Does action a still look best?

Optimal Decision Rule?

- **Two models** make predictions of **two different rewards** of a patient with state *s* under 4 different actions, a, a, a, and a.

Predicted $(R^{(0)}, R^{(1)})$ for patient with $S = s$



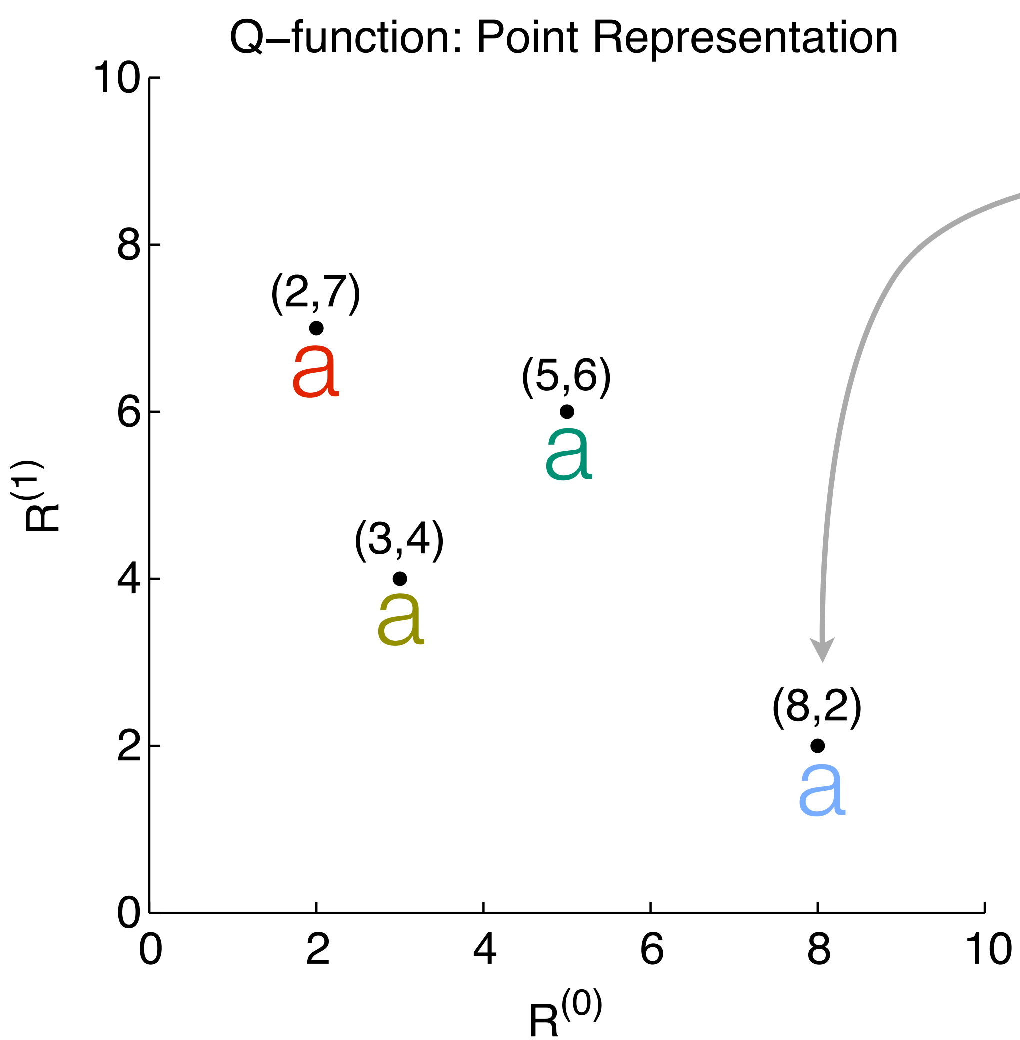
Does action a still look best?

Depends which reward matters most!

Optimal Decision Rule?

- **Two models** make predictions of **two different rewards** of a patient with state *s* under 4 different actions, a, a, a, and a.

Predicted ($R^{(0)}$, $R^{(1)}$) for patient with $S = s$



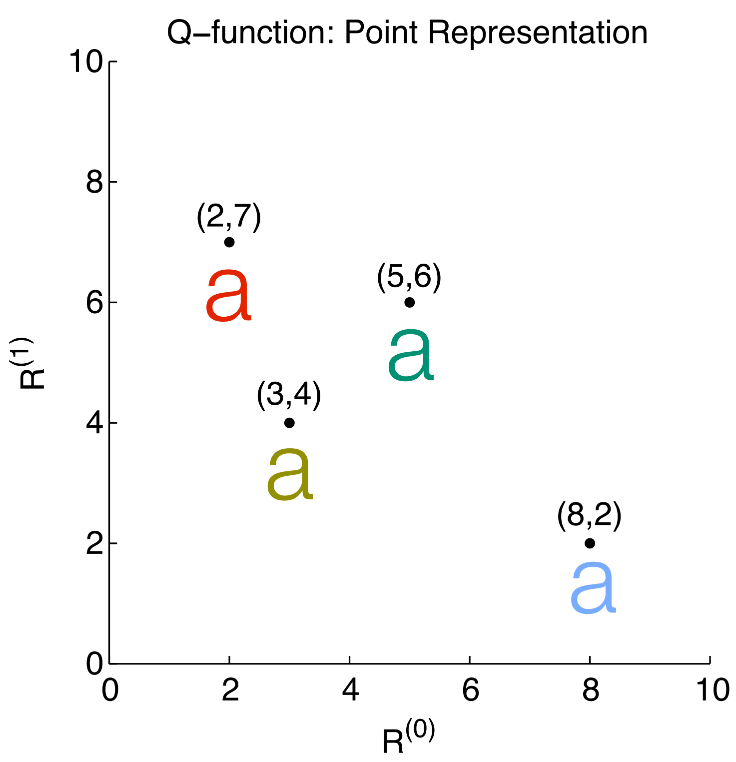
Does action a still look best?

Depends which reward matters most!

To decide, we need more information.

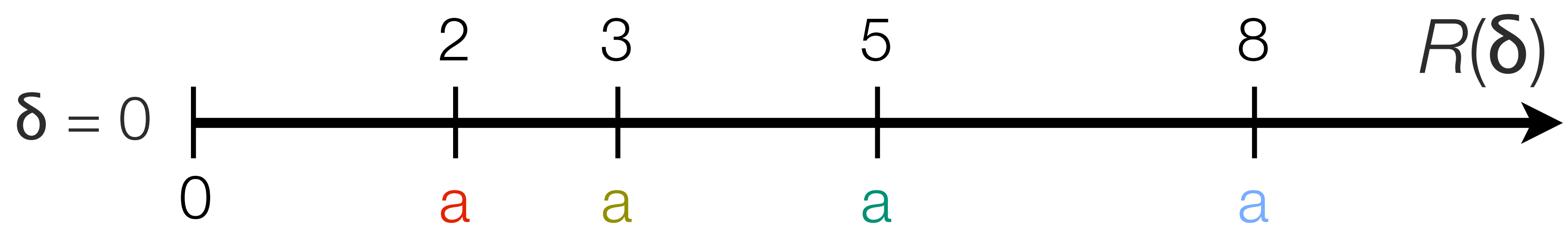
Mathematizing "Preference"

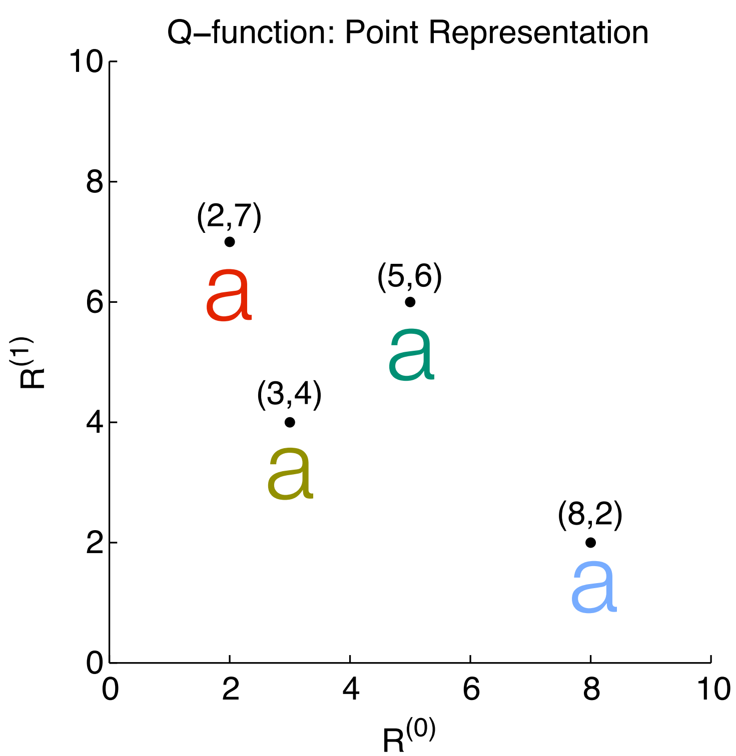
- Define a new reward $R(\delta) \equiv (1 - \delta) \cdot R^{(0)} + \delta \cdot R^{(1)}$
- $0 <= \delta <= 1$
- Proceed as before



Mathematizing “Preference”

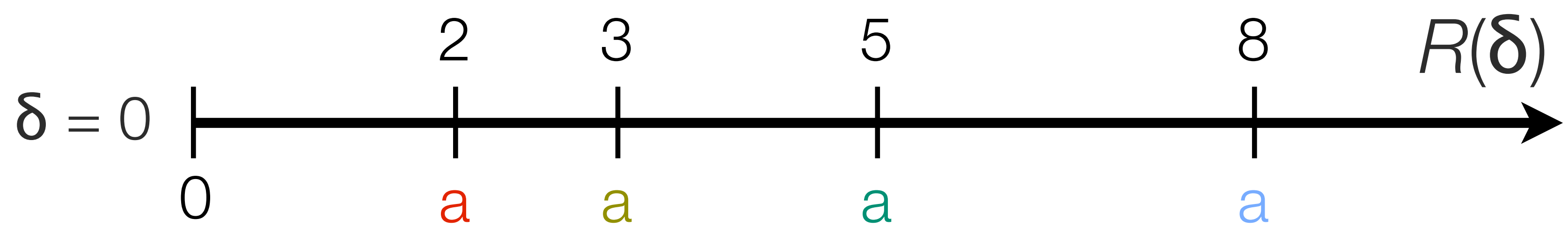
- Define a new reward $R(\delta) \equiv (1 - \delta) \cdot R^{(0)} + \delta \cdot R^{(1)}$
- $0 <= \delta <= 1$
- Proceed as before

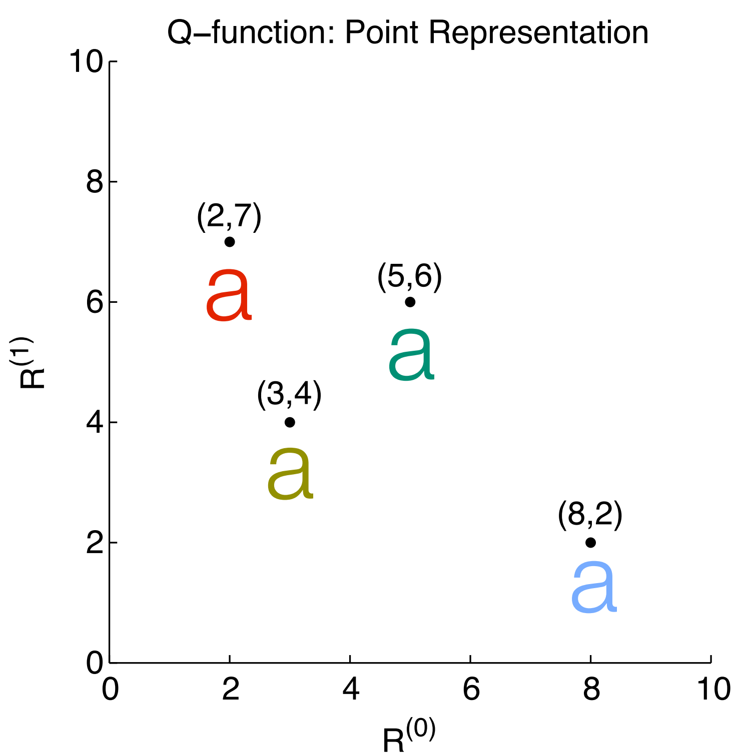


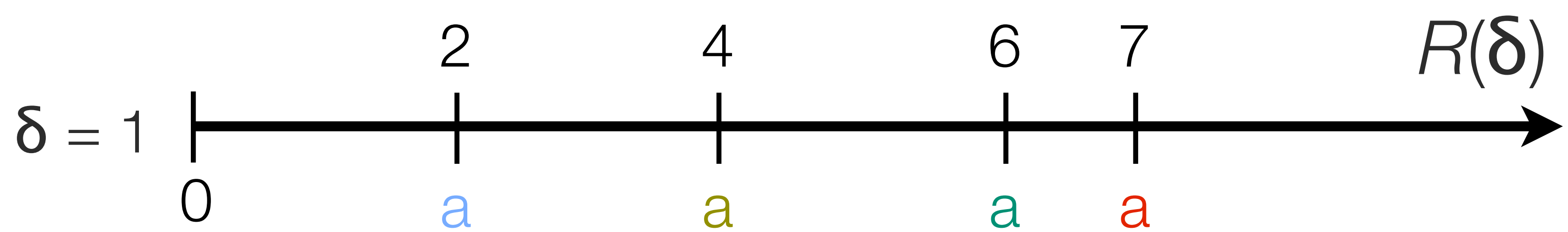


Mathematizing “Preference”

- Define a new reward $R(\boldsymbol{\delta}) \equiv (1 - \boldsymbol{\delta}) \cdot R^{(0)} + \boldsymbol{\delta} \cdot R^{(1)}$
- $0 <= \boldsymbol{\delta} <= 1$
- Proceed as before

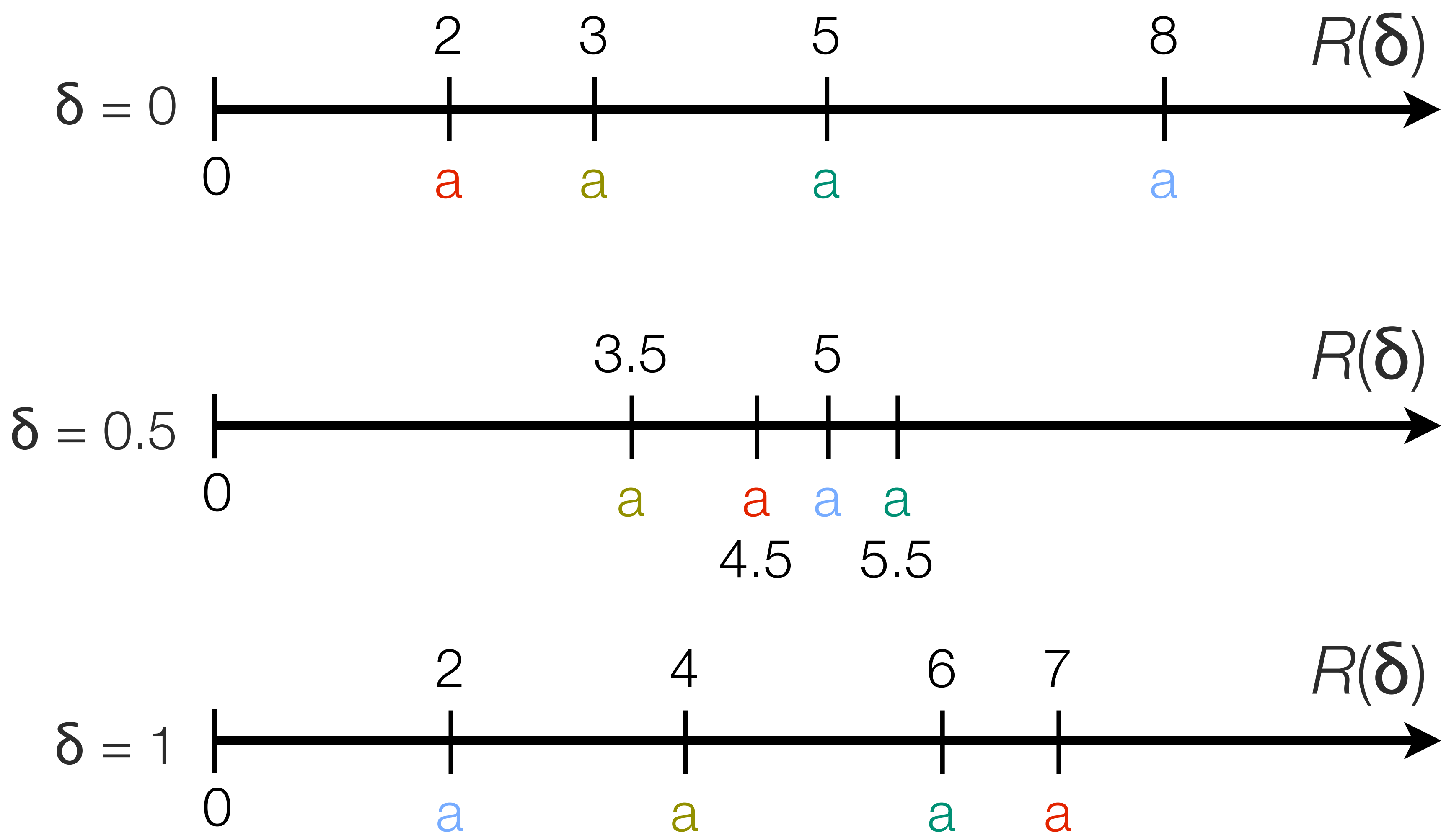


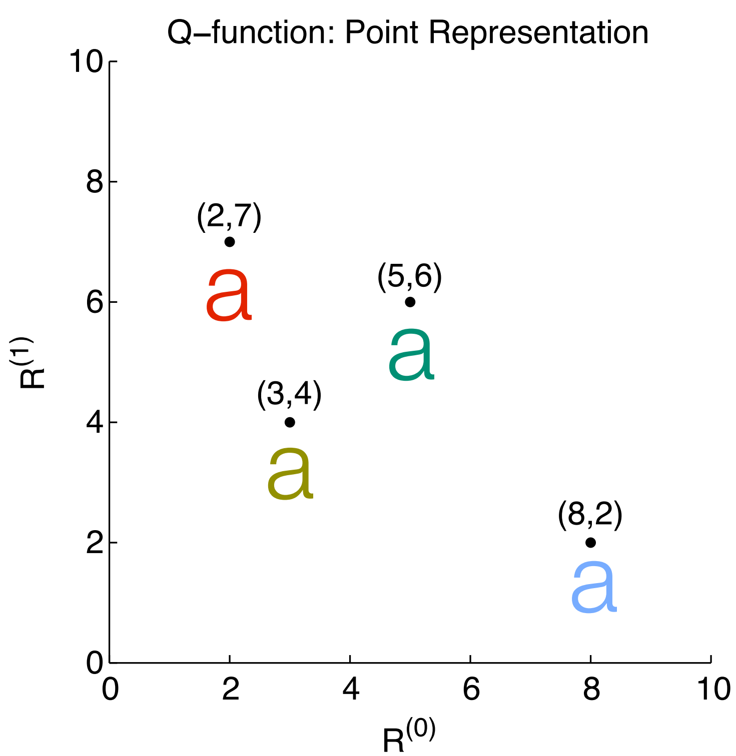




Mathematizing “Preference”

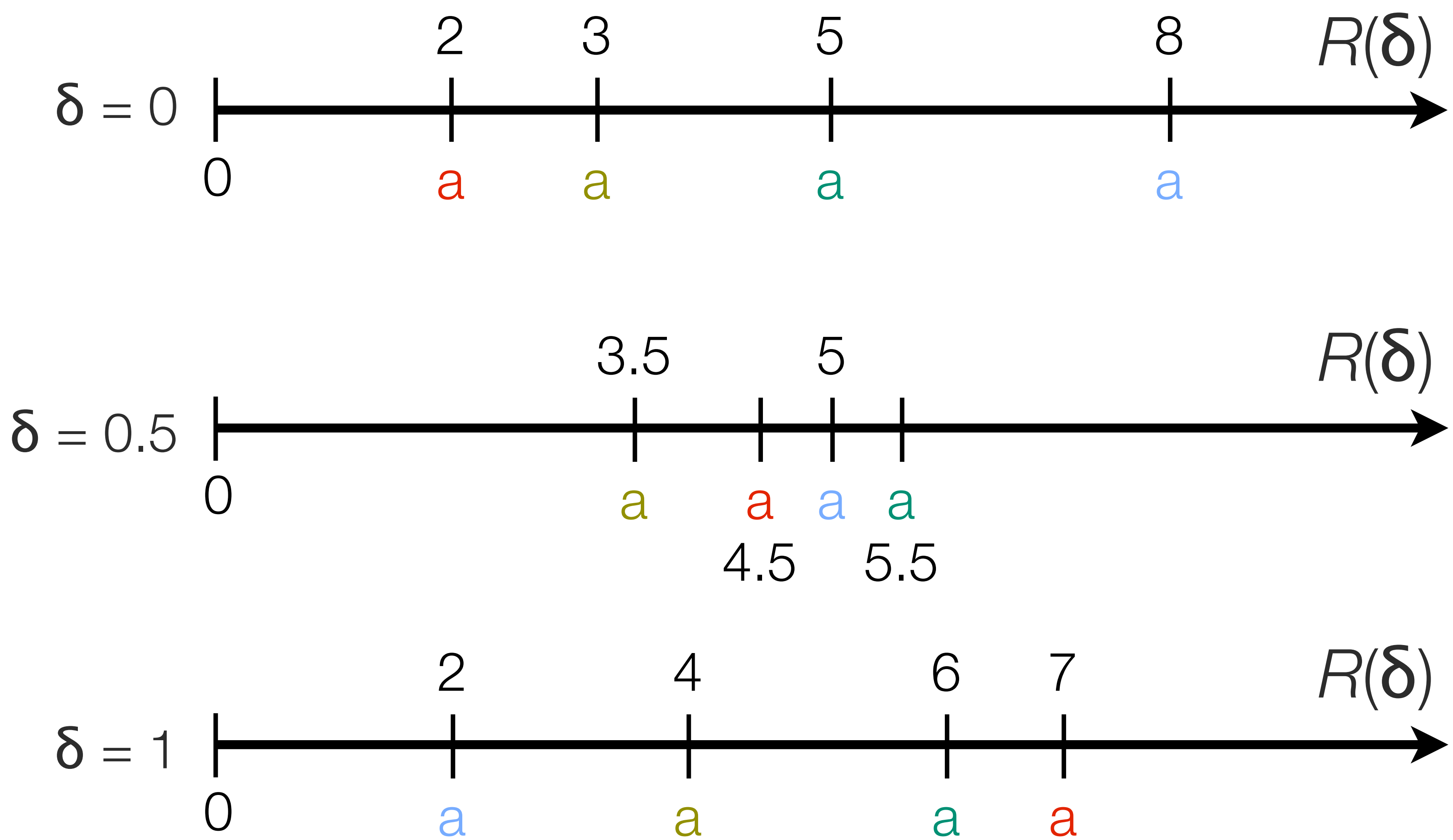
- Define a new reward $R(\delta) \equiv (1 - \delta)\cdot R^{(0)} + \delta \cdot R^{(1)}$
- $0 <= \delta <= 1$
- Proceed as before

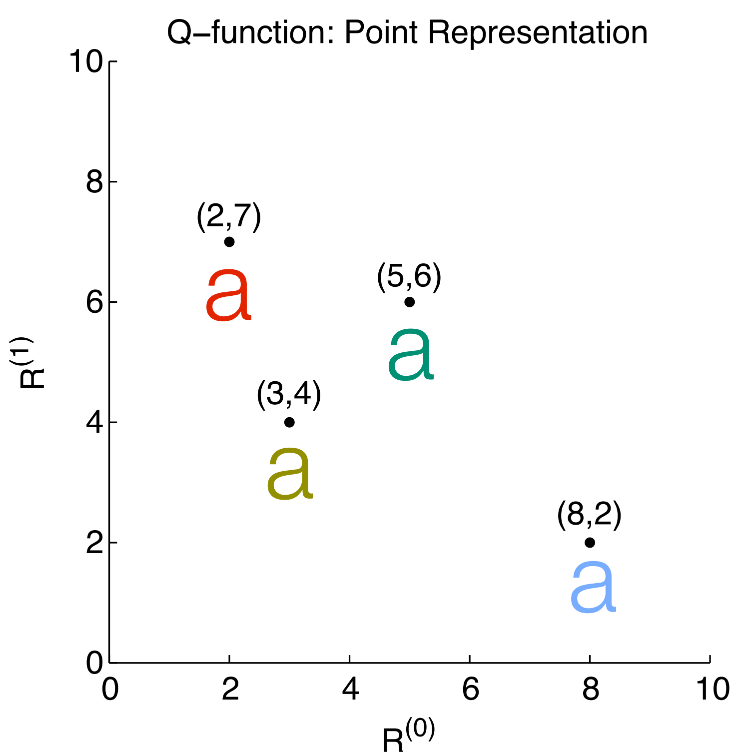




Mathematizing "Preference"

- Define a new reward $R(\delta) \equiv (1 - \delta) \cdot R^{(0)} + \delta \cdot R^{(1)}$
- $0 <= \delta <= 1$
- Proceed as before





- δ represents "How much do I care about $R^{(1)}$?"

Preference Elicitation

Preference Elicitation

- Intuitively, δ represents “How much do I care about $R^{(1)}$?”

Preference Elicitation

- Intuitively, δ represents “How much do I care about $R^{(1)}$?”
- Preference Elicitation Approach:

Preference Elicitation

- Intuitively, δ represents "How much do I care about $R^{(1)}$?"
- Preference Elicitation Approach:
  - Figure out the decision maker's δ

Preference Elicitation

- Intuitively, δ represents "How much do I care about $R^{(1)}$?"
- Preference Elicitation Approach:
  - Figure out the decision maker's δ
  - Define a reward $R(\delta) \equiv (1 - \delta) \cdot R^{(0)} + \delta \cdot R^{(1)}$

Preference Elicitation

- Intuitively, δ represents "How much do I care about $R^{(1)}$?"
- Preference Elicitation Approach:
  - Figure out the decision maker's δ
  - Define a reward $R(\delta) \equiv (1 - \delta) \cdot R^{(0)} + \delta \cdot R^{(1)}$
  - Construct the optimal decision rule for that reward

Preference Elicitation

- Intuitively, δ represents "How much do I care about $R^{(1)}$?"
- Preference Elicitation Approach:
  - Figure out the decision maker's δ
  - Define a reward $R(\delta) \equiv (1 - \delta) \cdot R^{(0)} + \delta \cdot R^{(1)}$
  - Construct the optimal decision rule for that reward
    - Construct predictor of $R(\delta)$ using regression

Preference Elicitation

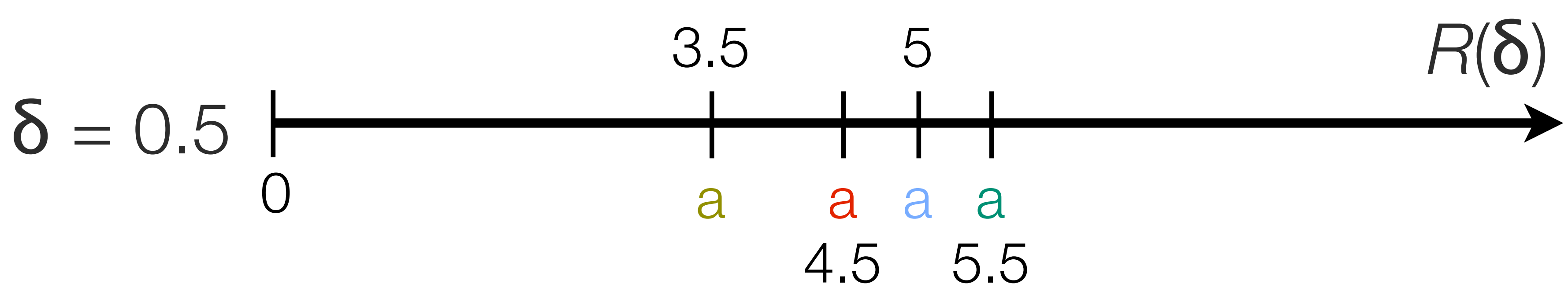
- Intuitively, δ represents "How much do I care about $R^{(1)}$?"
- Preference Elicitation Approach:
  - Figure out the decision maker's δ
  - Define a reward $R(\delta) \equiv (1 - \delta) \cdot R^{(0)} + \delta \cdot R^{(1)}$
  - Construct the optimal decision rule for that reward
    - Construct predictor of $R(\delta)$ using regression
    - Consider the $R(\delta)$ under different actions

Preference Elicitation

- Intuitively, δ represents "How much do I care about $R^{(1)}$?"
- Preference Elicitation Approach:
  - Figure out the decision maker's δ
  - Define a reward $R(\delta) \equiv (1 - \delta) \cdot R^{(0)} + \delta \cdot R^{(1)}$
  - Construct the optimal decision rule for that reward
    - Construct predictor of $R(\delta)$ using regression
    - Consider the $R(\delta)$ under different actions
    - Recommend the best one

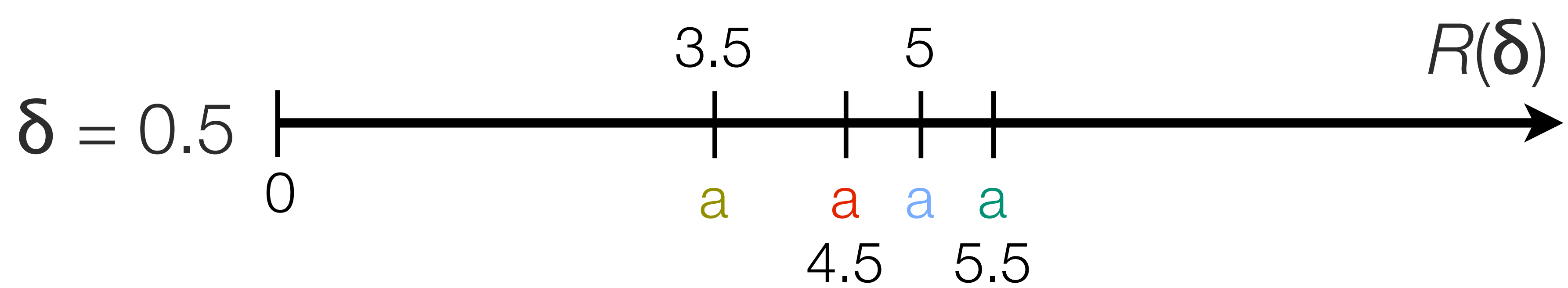
Preference Elicitation

- Intuitively, δ represents “How much do I care about $R^{(1)}$?”
- Preference Elicitation Approach:
  - Figure out the decision maker’s δ
  - Define a reward $R(\delta) \equiv (1 - \delta) \cdot R^{(0)} + \delta \cdot R^{(1)}$
  - Construct the optimal decision rule for that reward
    - Construct predictor of $R(\delta)$ using regression
    - Consider the $R(\delta)$ under different actions
    - Recommend the best one



Preference Elicitation

- Intuitively, δ represents "How much do I care about $R^{(1)}$?"
- Preference Elicitation Approach:
  - **Figure out the decision maker's δ**
  - Define a reward $R(\delta) \equiv (1 - \delta) \cdot R^{(0)} + \delta \cdot R^{(1)}$
  - Construct the optimal decision rule for that reward
    - Construct predictor of $R(\delta)$ using regression
    - Consider the $R(\delta)$ under different actions
    - Recommend the best one



Preference Elicitation

Preference Elicitation

- Figure out the decision maker's δ

Preference Elicitation

- Figure out the decision maker's δ

- "Consider two actions.
  You can have (8, 5), or you can have (5, *x*).
  What value of *x* makes you indifferent to this choice?"*

*Actual questioning would be much more subtle.

Preference Elicitation

- Figure out the decision maker's δ
- "Consider two actions.
  You can have (8, 5), or you can have (5, *x*).
  What value of *x* makes you indifferent to this choice?"*
- Find δ so that $R(\delta)$ is equal for the two points
  - $(1-\delta)\cdot 8 + \delta\cdot 5 = (1-\delta)\cdot 4 + \delta\cdot x$

*Actual questioning would be much more subtle.

Preference Elicitation

- Figure out the decision maker's δ
- "Consider two actions.
  You can have (8, 5), or you can have (5, *x*).
  What value of *x* makes you indifferent to this choice?"*
- Find δ so that $R(\delta)$ is equal for the two points
  - $(1-\delta)\cdot 8 + \delta\cdot 5 = (1-\delta)\cdot 4 + \delta\cdot x$
- **Note that this approach does not have anything to do with the actions that are actually available.**

*Actual questioning would be much more subtle.

Inverse Preference Elicitation

Inverse Preference Elicitation

- There are an infinite number of δ, but there are only 4 actions.

Inverse Preference Elicitation

- There are an infinite number of δ, but there are only 4 actions.
- Preference Elicitation
  - "Give me your δ, I will tell you the right action."

Inverse Preference Elicitation

- There are an infinite number of δ, but there are only 4 actions.
- Preference Elicitation
    - "Give me your δ, I will tell you the right action."
- **Inverse Preference Elicitation**
    - *"Give me your action, I will tell you your* δ*."*

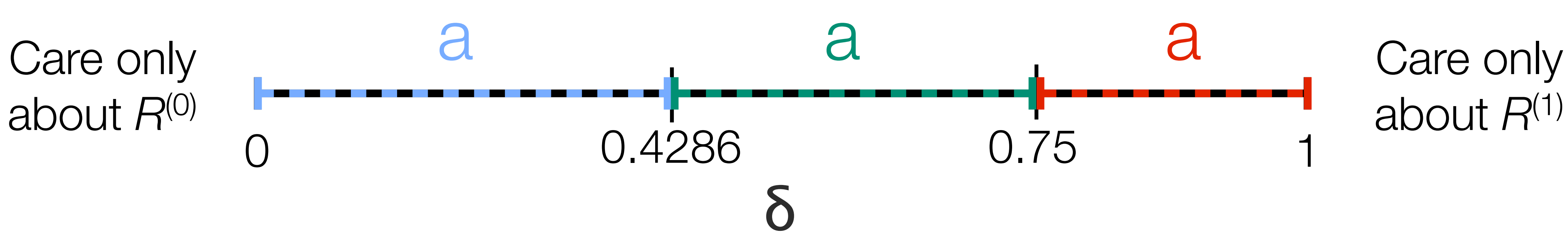
Inverse Preference Elicitation

Inverse Preference Elicitation

- *"Give me your action, I will tell you your* δ*."*
    - In fact, each action is optimal over a range of δ

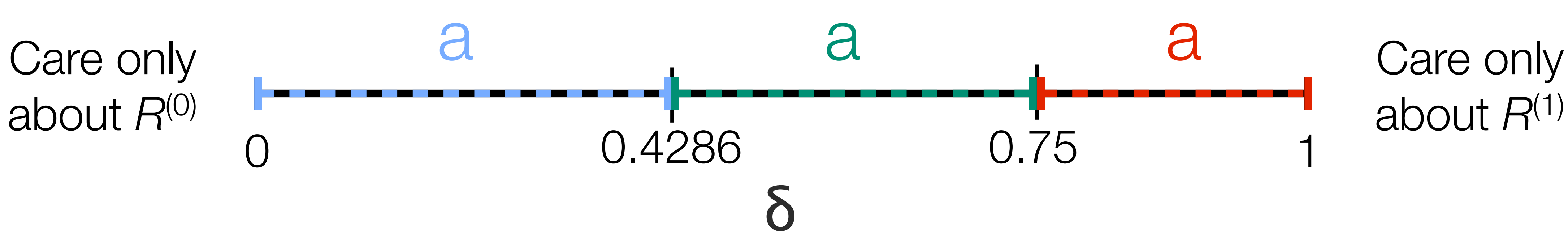
Inverse Preference Elicitation

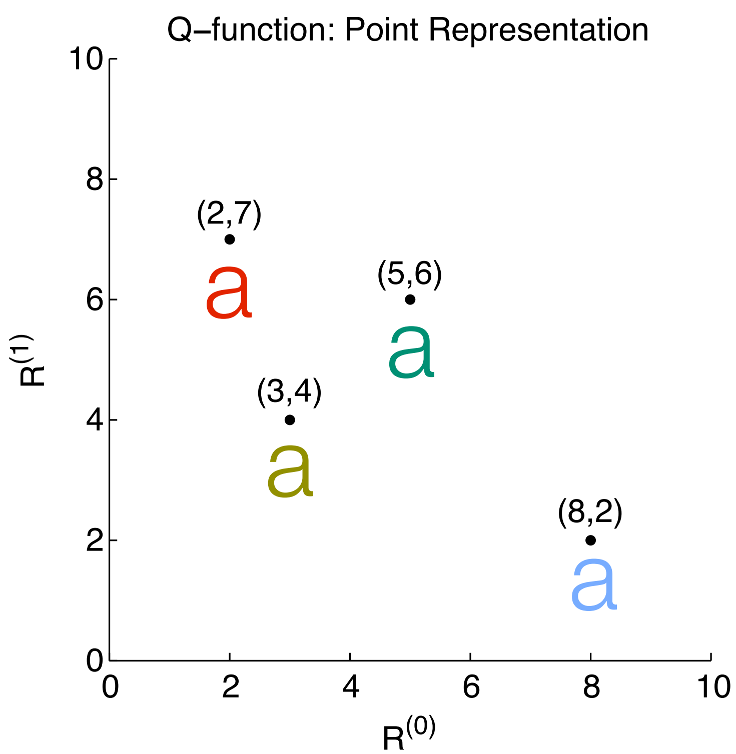
- *"Give me your action, I will tell you your* δ*."*
  - In fact, each action is optimal over a range of δ



Inverse Preference Elicitation

- *"Give me your action, I will tell you your* δ*."*
  - In fact, each action is optimal over a range of δ





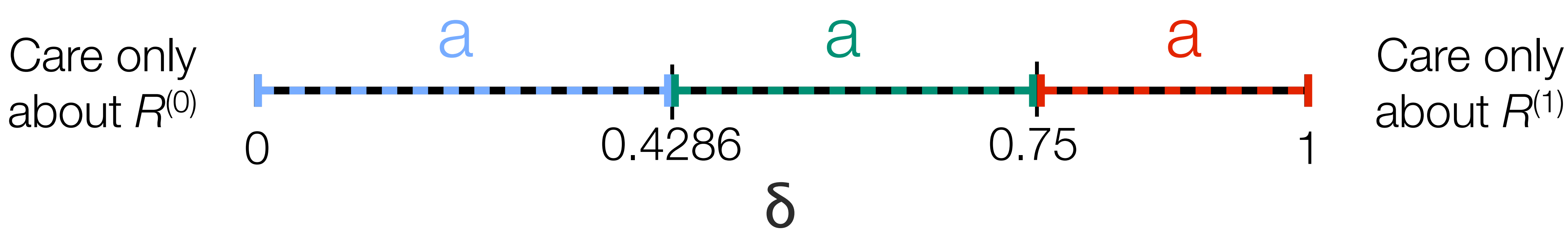
- Note that a does not appear anywhere.

Possible Decision Aid

Care only about $R^{(0)}$

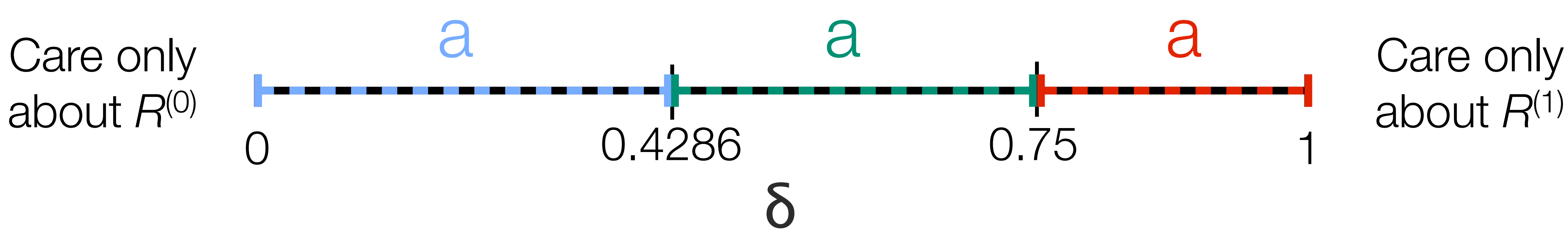
Care only about $R^{(1)}$

Possible Decision Aid



- May want to coarsen the above information:

Possible Decision Aid



- May want to coarsen the above information:

"I am concerned..."

Mostly about $R^{(0)}$	Somewhat more about $R^{(0)}$ than $R^{(1)}$	Somewhat more about $R^{(1)}$ than $R^{(0)}$	Mostly about $R^{(1)}$
a	a or a	a	a

Preference Elicitation vs. Inverse Preference Elicitation

- Inverse Preference Elicitation
  - Method for choosing an action when faced with multiple rewards
  - Provides information about available actions
  - Choice among a **small number** of alternatives
- Preference Elicitation
  - Method for choosing an action when faced with multiple rewards
  - Provides **no** information about available actions
  - Choice among an **infinite** number of alternatives

"I am concerned..."

Mostly about $R^{(0)}$	Somewhat more about $R^{(0)}$ than $R^{(1)}$	Somewhat more about $R^{(1)}$ than $R^{(0)}$	Mostly about $R^{(1)}$
a	a or a	a	a

End Part I

- We’ve covered:
    - Optimal Decision Rules
    - Mathematizing Preference
    - Preference Elicitation
    - Inverse Preference Elicitation

- Pause for questions and discussion?

Symptoms and Side-Effects in Schizophrenia
Sequences of Treatment

Symptoms and Side-Effects in Schizophrenia **Sequences** of Treatment

- Many treatments available for treating schizophrenia (dozens)
- Evidence-based medicine would look at predicted outcomes, recommend a ***sequence of actions***: schizophrenia is chronic
- *At least* two important objectives:
    - Maximize symptom reduction, minimize weight gain

Symptoms and Side-Effects in Schizophrenia **Sequences** of Treatment

- Many treatments available for treating schizophrenia (dozens)
- Evidence-based medicine would look at predicted outcomes, recommend a ***sequence of actions***: schizophrenia is chronic
- *At least* two important objectives:
    - Maximize symptom reduction, minimize weight gain
- Treatments that provide the best symptom reduction induce the worst weight gain, and vice-versa

Symptoms and Side-Effects in Schizophrenia **Sequences** of Treatment

- Many treatments available for treating schizophrenia (dozens)
- Evidence-based medicine would look at predicted outcomes, recommend a ***sequence of actions***: schizophrenia is chronic
- *At least* two important objectives:
    - Maximize symptom reduction, minimize weight gain
- Treatments that provide the best symptom reduction induce the worst weight gain, and vice-versa
- Different doctors and patients have very different preferences about relative importance of outcomes

Symptoms and Side-Effects in Schizophrenia **Sequences** of Treatment

- Many treatments available for treating schizophrenia (dozens)
- Evidence-based medicine would look at predicted outcomes, recommend a ***sequence of actions***: schizophrenia is chronic
- *At least* two important objectives:
    - Maximize symptom reduction, minimize weight gain
- Treatments that provide the best symptom reduction induce the worst weight gain, and vice-versa
- Different doctors and patients have very different preferences about relative importance of outcomes
    - **How can we recommend a *sequence of actions* that accommodates these preferences?**

Learning a Sequence of Actions From Data

Learning a Sequence of Actions From Data

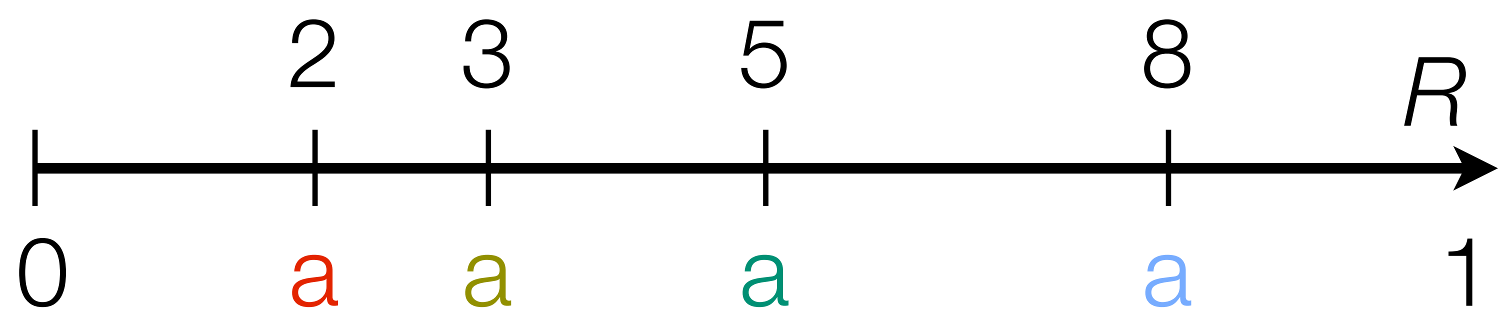
- (S_1, A_1, S_2, A_2, R) for each individual
  - S_j - "State" - Patient covariates (previous txts, response,...)
  - A_j - "Action" - Treatment offered to the patient
  - R - "Reward" - Clinical outcome
- Actions A_j have known randomization probability

Learning a Sequence of Actions From Data

- (S_1,A_1,S_2,A_2,R) for each individual
  - S_j - "State" - Patient covariates (previous txts, response,...)
  - A_j - "Action" - Treatment offered to the patient
  - R - "Reward" - Clinical outcome
- Actions A_j have known randomization probability
- Let's start by looking at Stage 2: (S_2,A_2,R)

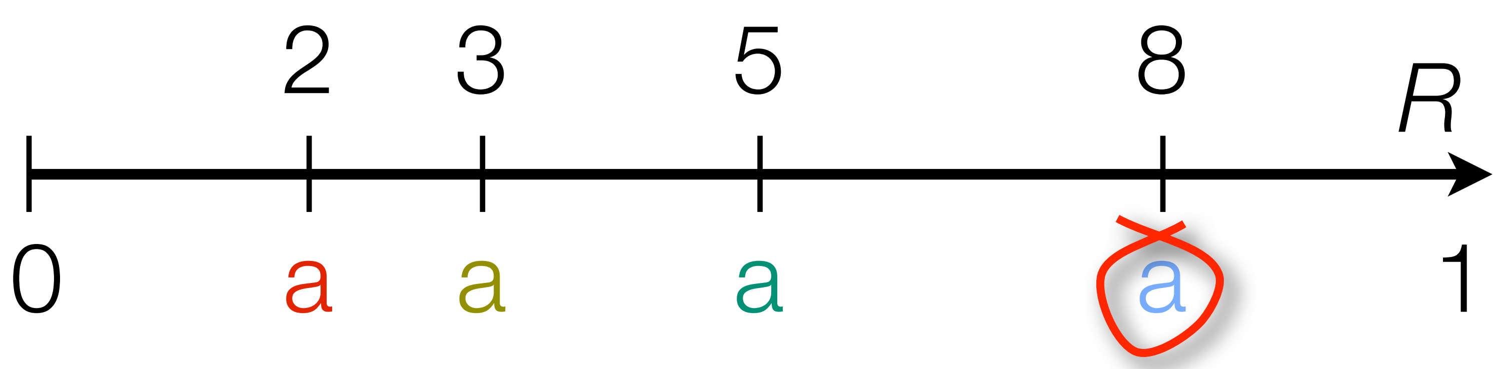
Stage 2: Just like the 1-stage case

- A model makes predictions of the reward of a patient with state $S_2{=}s_2$ under 4 different actions, a, a, a, and a.
- For the patient with the predictions shown below, action a looks best. (Higher rewards are better.)



Stage 2: Just like the 1-stage case

- A model makes predictions of the reward of a patient with state $S_2 = s_2$ under 4 different actions, a, a, a, and a.
- For the patient with the predictions shown below, action a looks best. (Higher rewards are better.)

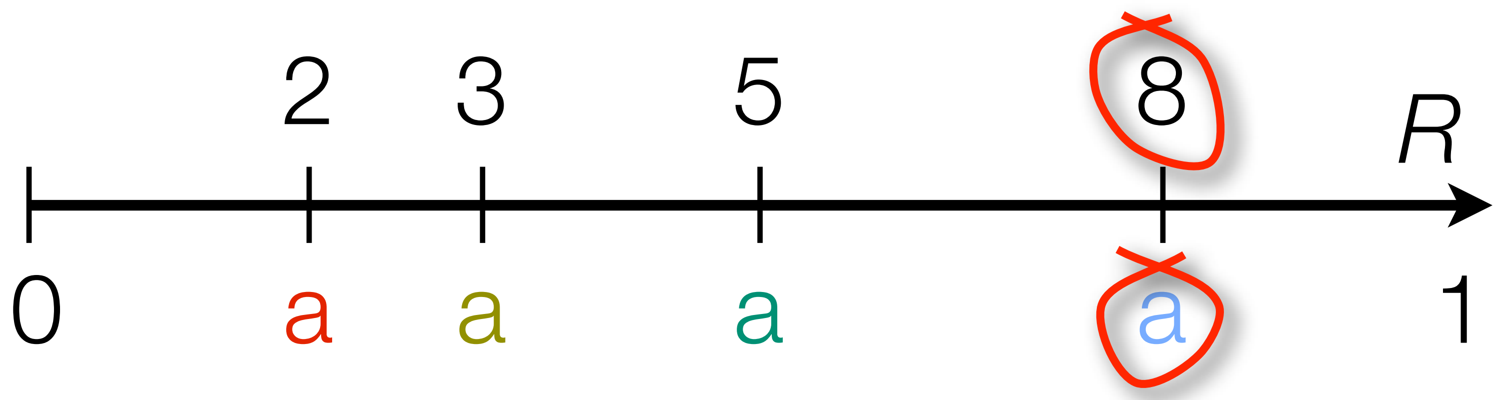


We know what action we would take

Stage 2: Just like the 1-stage case

- A model makes predictions of the reward of a patient with state $S_2=s_2$ under 4 different actions, a, a, a, and a.
- For the patient with the predictions shown below, action a looks best. (Higher rewards are better.)

We can predict the reward R we will get

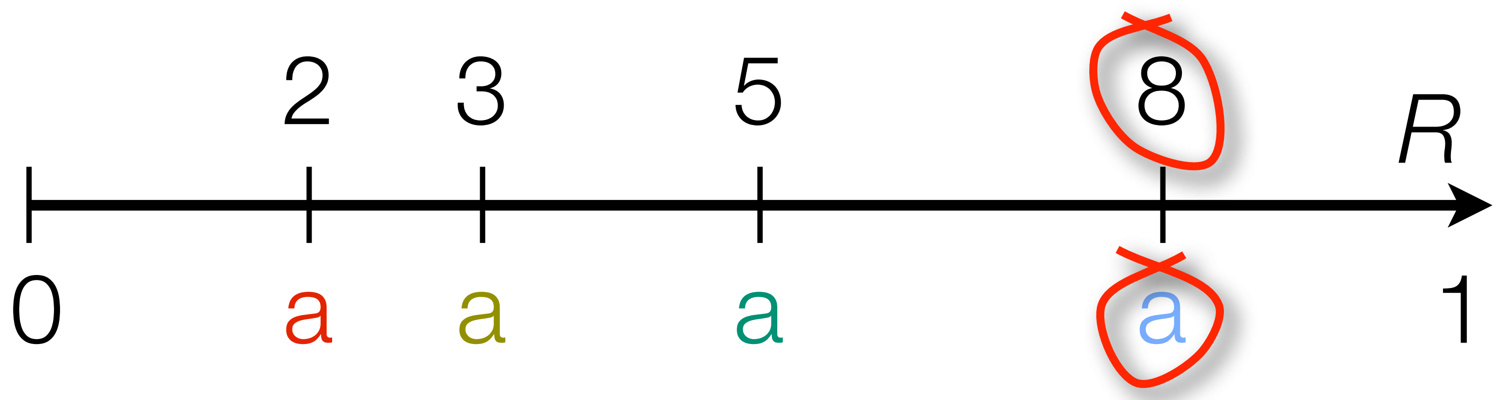


We know what action we would take

Stage 2: Just like the 1-stage case

- A model makes predictions of the reward of a patient with state $S_2 = s_2$ under 4 different actions, a, a, a, and a.
- For the patient with the predictions shown below, action a looks best. (Higher rewards are better.)

We can predict the reward R we will get



We know what action we would take

We now have our decision rule for Stage 2

(we have to do the above for all s_2)

Stage 1: Dynamic Programming

Stage 1: Dynamic Programming

- Recall: We have (S_1,A_1,S_2,A_2,R)

Stage 1: Dynamic Programming

- Recall: We have (S_1,A_1,S_2,A_2,R)
- How do we choose A_1, i.e., how do we make a decision rule?

Stage 1: Dynamic Programming

- Recall: We have (S_1,A_1,S_2,A_2,R)
- How do we choose A_1, i.e., how do we make a decision rule?
- We now know, for any s_2:

Stage 1: Dynamic Programming

- Recall: We have (S_1,A_1,S_2,A_2,R)
- How do we choose A_1, i.e., how do we make a decision rule?
- We now know, for any s_2:
    - The action we should take

Stage 1: Dynamic Programming

- Recall: We have (S_1,A_1,S_2,A_2,R)
- How do we choose A_1, i.e., how do we make a decision rule?
- We now know, for any s_2:
  - The action we should take
  - The expected reward we would get (really, a prediction of this)

Stage 1: Dynamic Programming

- Recall: We have (S_1,A_1,S_2,A_2,R)
- How do we choose A_1, i.e., how do we make a decision rule?
- We now know, for any s_2:
  - The action we should take
  - The expected reward we would get (really, a prediction of this)
- Let's construct a decision rule for A_1
  assuming we follow our optimal rule for stage 2.

Stage 1: Dynamic Programming

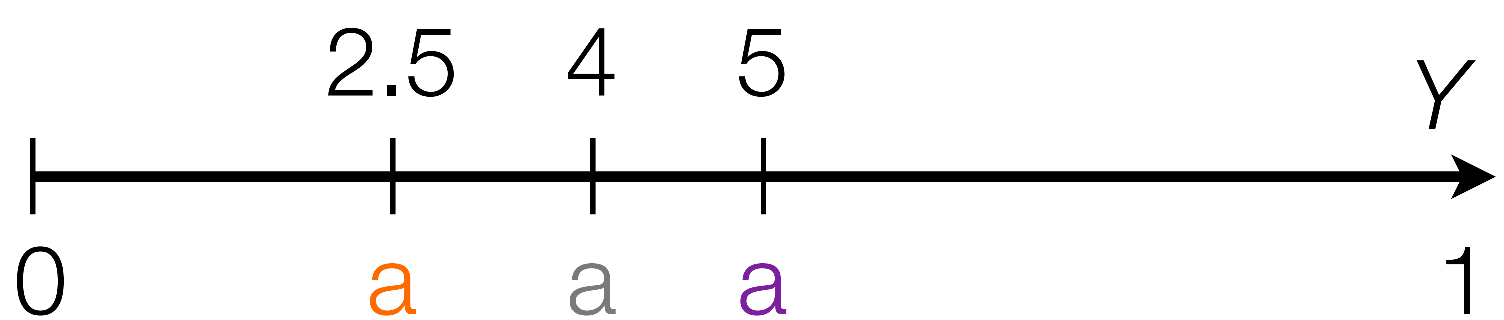
- Recall: We have (S_1,A_1,S_2,A_2,R)
- How do we choose A_1, i.e., how do we make a decision rule?
- We now know, for any s_2:
  - The action we should take
  - The expected reward we would get (really, a prediction of this)
- Let's construct a decision rule for A_1
  assuming we follow our optimal rule for stage 2.
  - **Create** a dataset (S_1,A_1,Y) where Y is the predicted optimal reward for the s_2 in the original data.

Stage 1: Dynamic Programming

- Recall: We have (S_1,A_1,S_2,A_2,R)
- How do we choose A_1, i.e., how do we make a decision rule?
- We now know, for any s_2:
  - The action we should take
  - The expected reward we would get (really, a prediction of this)
- Let's construct a decision rule for A_1
  assuming we follow our optimal rule for stage 2.
  - **Create** a dataset (S_1,A_1,Y) where Y is the predicted optimal reward for the s_2 in the original data.
  - Regress Y on S_1,A_1

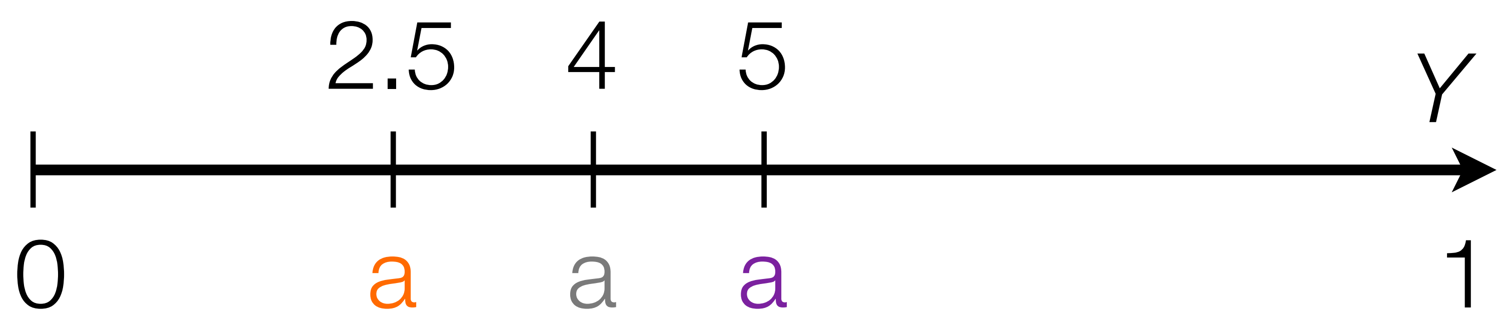
Stage 1

- A model makes predictions of the “pseudo-outcome” Y of a patient with state $S_1 = s_1$ under 3 different actions, a, a, and a.
  - Predicts reward of patient if we choose a and then act optimally
- For the patient shown below, action a looks best. (Higher pseudo-outcomes are better.)



Stage 1

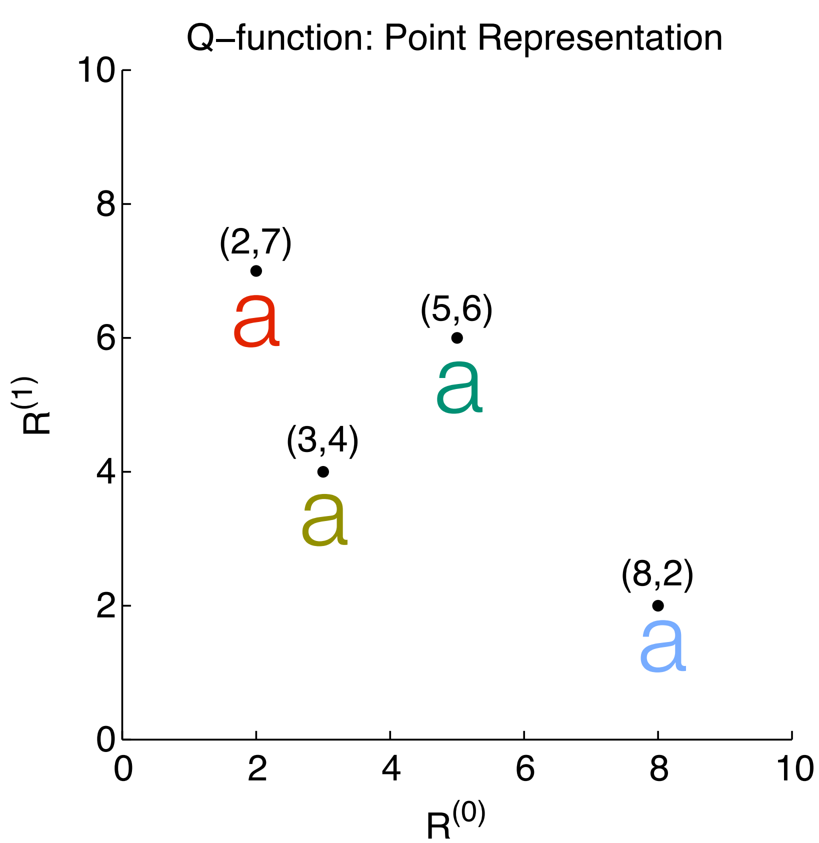
- A model makes predictions of the "pseudo-outcome" Y of a patient with state $S_1=s_1$ under 3 different actions, a, a, and a.
  - Predicts reward of patient if we choose a and then act optimally
- For the patient shown below, action a looks best. (Higher pseudo-outcomes are better.)



We now have our decision rule for Stage 1

Dynamic Programming: Multiple Rewards

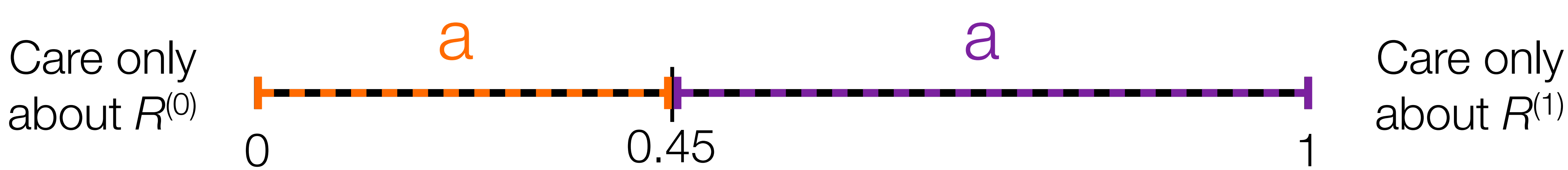
- Big "trick" was constructing *Y*
    - Requires knowing the decision rule at stage 2
    - But what if we don't know?



- We can still use the δ approach to make a single reward *R*(δ) and proceed as before

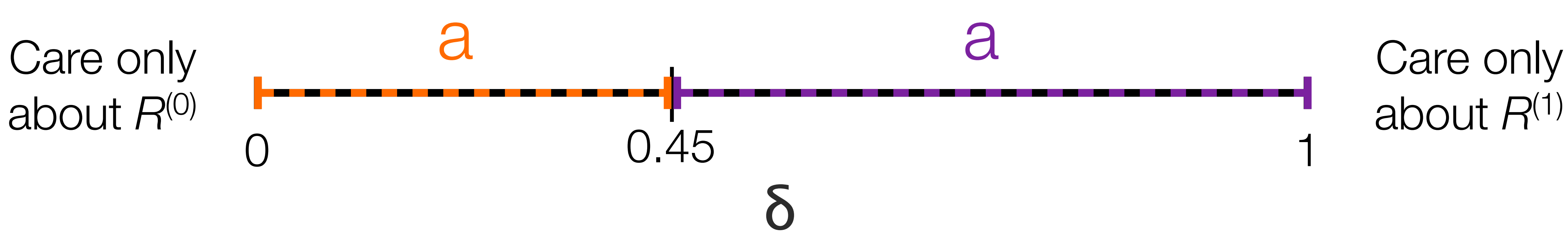
Dynamic Programming: Inverse Preference Elicitation

- We can still use the δ approach to make a single reward $R(\delta)$ and proceed as before.
- And we can do Inverse Preference Elicitation! Algorithm is complex. [Lizotte, Bowling, Murphy 2010]



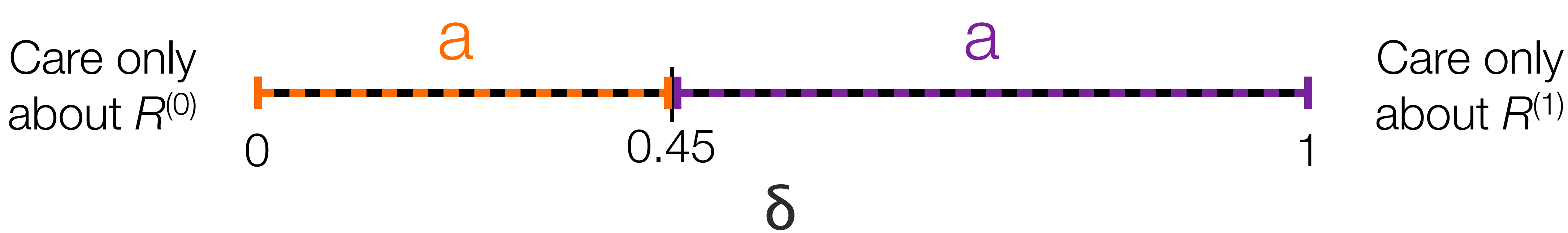
Dynamic Programming: Inverse Preference Elicitation

- We can still use the δ approach to make a single reward $R(\delta)$ and proceed as before.
- And we can do Inverse Preference Elicitation! Algorithm is complex. [Lizotte, Bowling, Murphy 2010]



Dynamic Programming: Inverse Preference Elicitation

- We can still use the δ approach to make a single reward $R(\delta)$ and proceed as before.
- And we can do Inverse Preference Elicitation! Algorithm is complex. [Lizotte, Bowling, Murphy 2010]



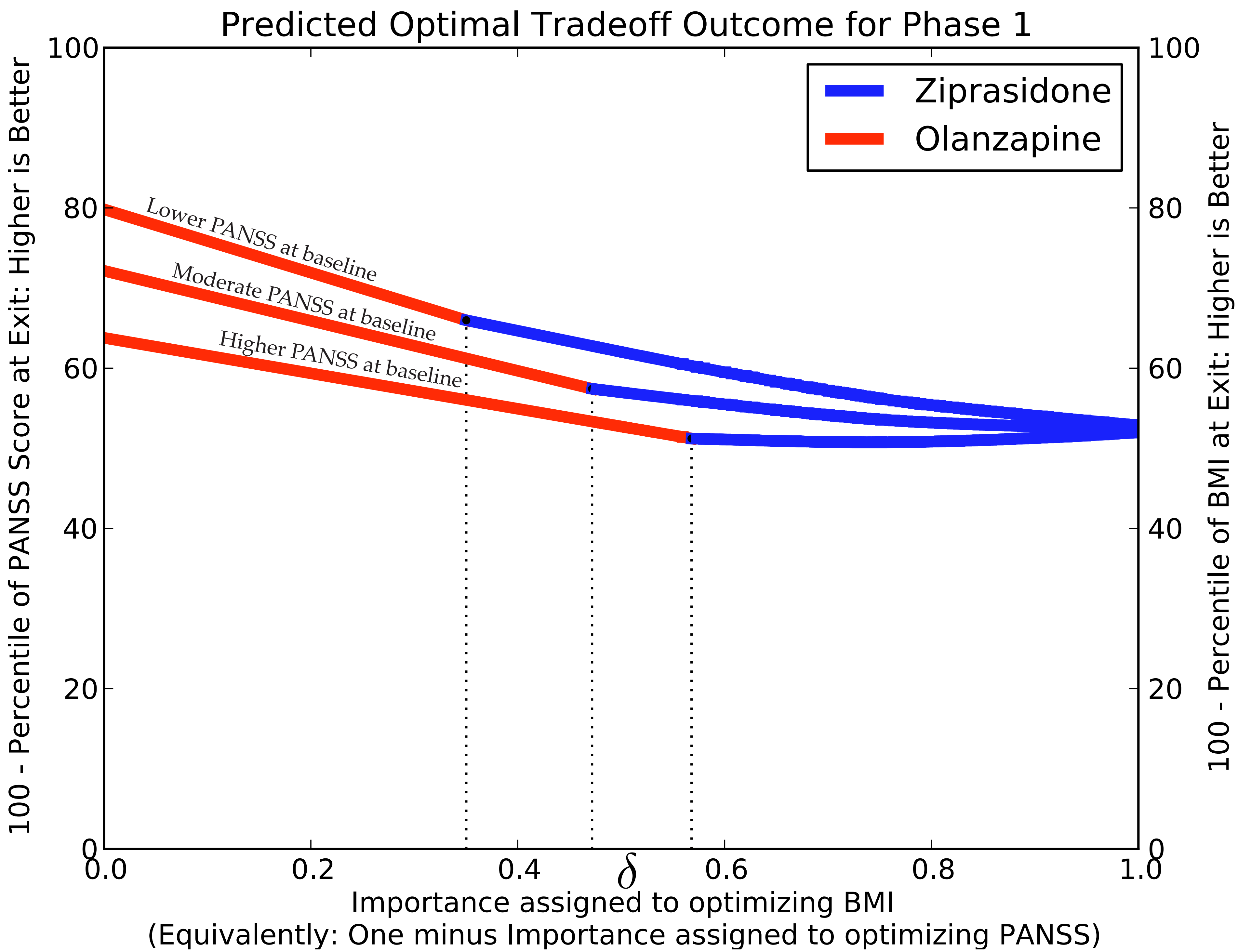
"I am concerned..."

Mostly about $R^{(0)}$	Somewhat more about $R^{(0)}$ than $R^{(1)}$	Somewhat more about $R^{(1)}$ than $R^{(0)}$	Mostly about $R^{(1)}$
a	a or a	a	a

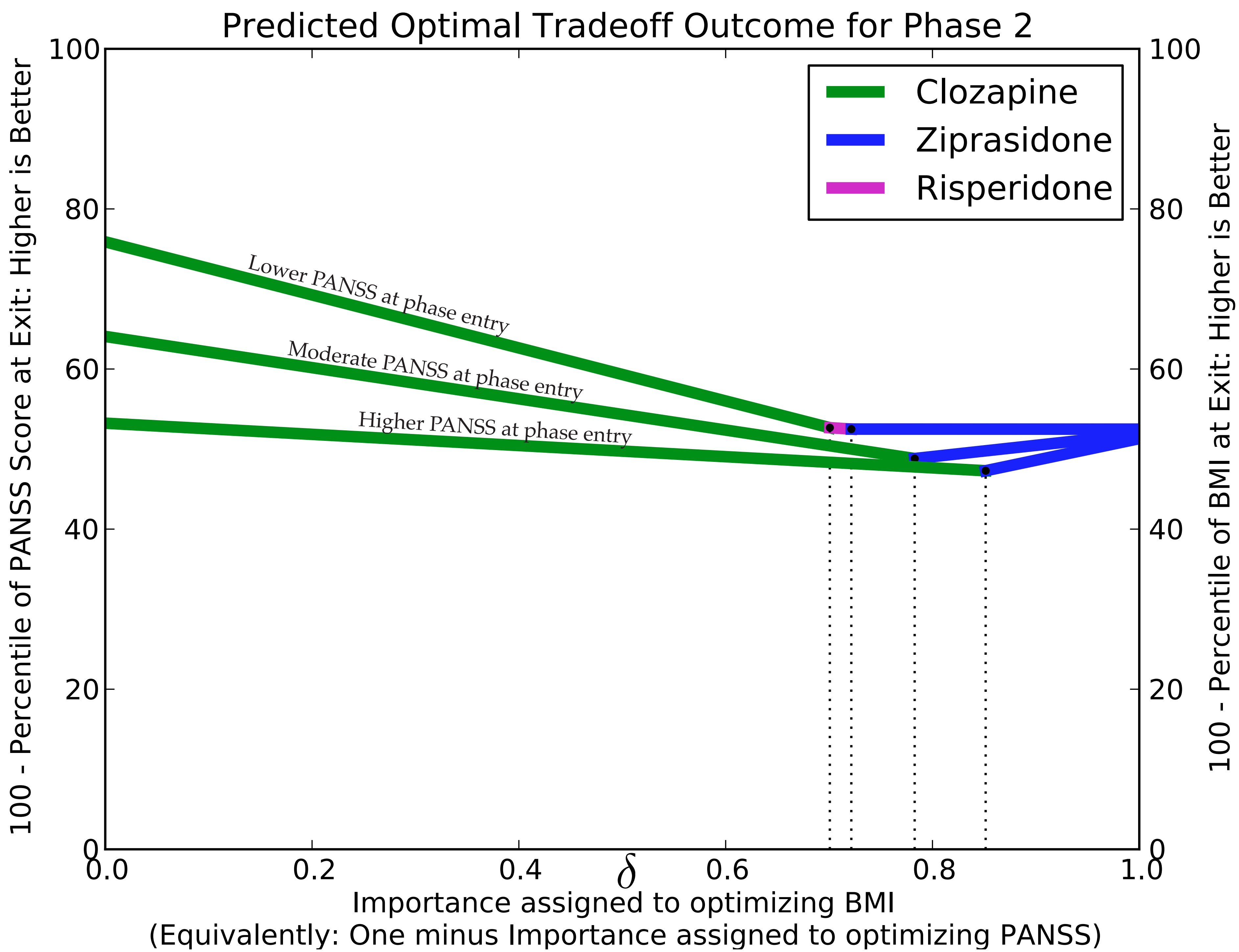
Example: CATIE

- Large (n = 1460) comparative effectiveness trial funded by NIMH
- Compares medications for treatment of schizophrenia
- Most patients randomized two times:
    - First to one of 5 actions
    - Then, if desired, to one of 5 different actions
- Details are quite complicated
- Following is a *highly* simplified analysis
- Overall, the results are consistent with what is known in the literature
- Rewards: PANSS (symptoms) versus BMI (weight gain side-effect)

Example: CATIE Exploratory Analysis



Example: CATIE Exploratory Analysis



Example: CATIE-based Decision Aid

- One possibility for a decision aid is a very coarse version of the plots:

Recommendation given State and Preference	Strong Preference for Symptom Relief over Weight Control	Mild Preference for Symptom Relief over Weight Control	Mild Preference for Weight Control over Symptom Relief	Strong Preference for Weight Control over Symptom Relief
Lower PANSS at Entry to Phase 1	Olanzapine	Olanzapine or Ziprasidone	Ziprasidone	Ziprasidone
Moderate PANSS at Entry to Phase 1	Olanzapine	Olanzapine or Ziprasidone	Ziprasidone	Ziprasidone
Higher PANSS at Entry to Phase 1	Olanzapine	Olanzapine	Olanzapine or Ziprasidone	Ziprasidone
Lower PANSS at Entry to Phase 2	Clozapine	Clozapine	Clozapine, Risperidone, or Ziprasidone	Ziprasidone
Moderate PANSS at Entry to Phase 2	Clozapine	Clozapine	Clozapine	Clozapine or Ziprasidone
Higher PANSS at Entry to Phase 2	Clozapine	Clozapine	Clozapine	Clozapine or Ziprasidone

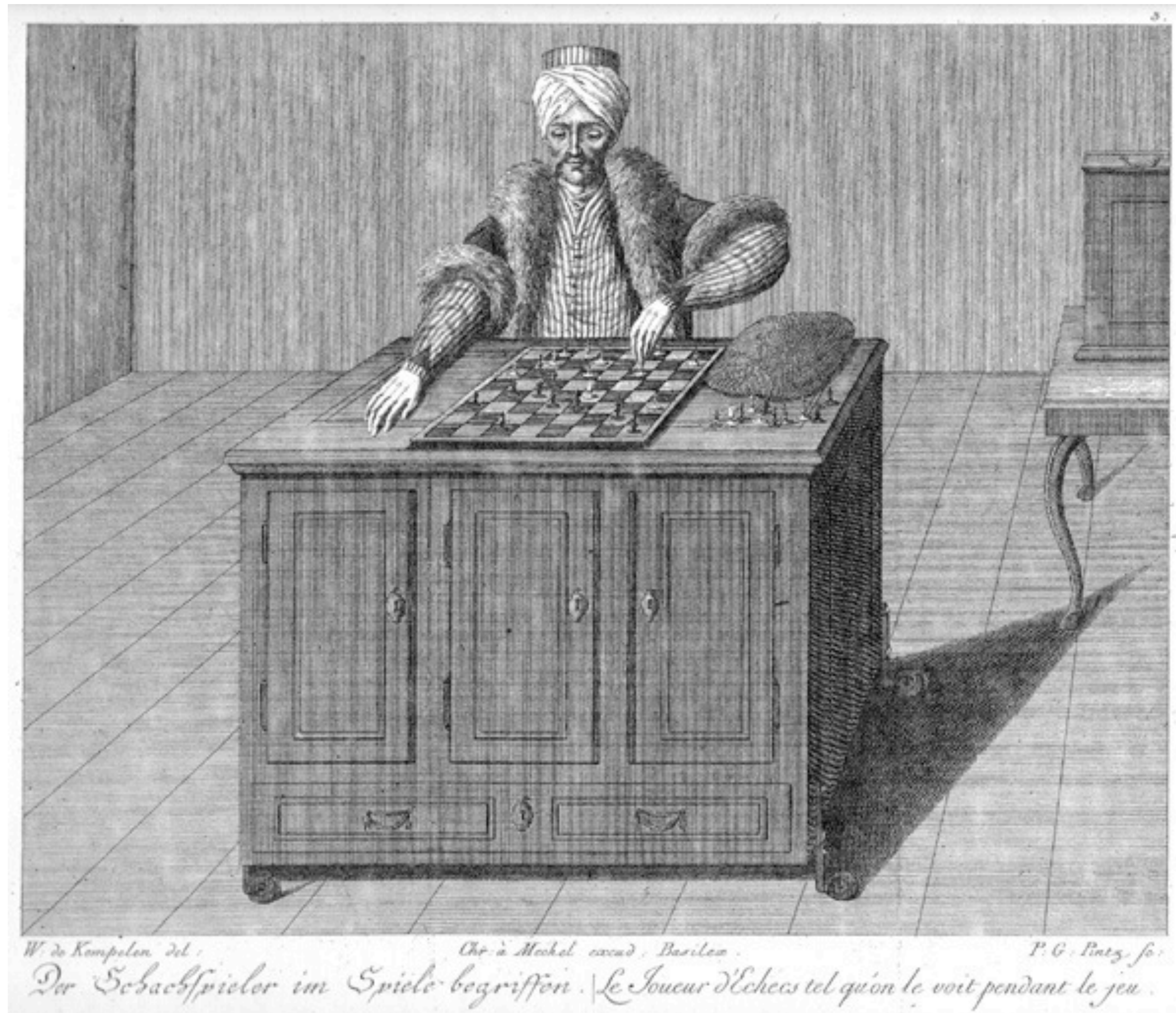
- Thanks to: Holly Wittemann, Brian Zikmund-Fisher for this idea

Future Work

- Evaluating the “Inverse Preference Elicitation” Idea
  - **MTurk Evaluation**
- The Algorithms and Methods
  - Measures of Uncertainty
  - More flexible models / Approximation algorithms
  - More reward definitions
- **Clinical Science Applications**

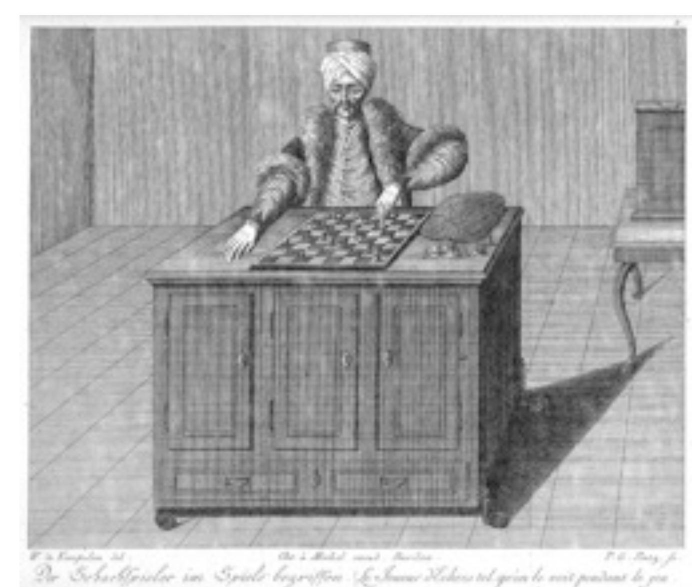
Amazon Mechanical Turk

- Mechanism for recruiting and paying users to do “Human Intelligence Tasks” - HITs
- Popular for running survey experiments (demographics at least as good as undergrads [Paolacci, Chandler, Ipeirotis 2010])



Amazon Mechanical Turk

- Our experiment will compare **eliciting δ using a slider** with directly **eliciting an action using a decision aid**.
- User will perform one of four different (similar and boring) sub-tasks, each one with different rate of pay and time duration
- The choice of action determines the sub-task, *and also* affects the workload of all the subsequent subtasks - myopic decision making is sub-optimal.
- Competing preferences:
    - Save time vs. Make money
- We will compare the appeal of the two methods
- Pilot in progress now



Future Work - Clinical Science

1. Schizophrenia
   - Symptom reduction versus functionality, or weight gain
2. Major Depressive Disorder
   - Symptom reduction versus weight gain, other side-effects
3. Type 2 Diabetes
   - Future disease complications versus drug side-effects

Questions

- Supported by National Institute of Health grants R01 MH080015 and P50 DA10075

- Daniel J. Lizotte, Michael Bowling, and Susan A. Murphy. *Efficient Reinforcement Learning with Multiple Reward Functions for Randomized Clinical Trial Analysis*. Proceedings of the Twenty-Seventh International Conference on Machine Learning (ICML), 2010.

- Related work:

  Barrett, L. and Narayanan, S. *Learning all optimal policies with multiple criteria.* In Proceedings of the 25th International Conference on Machine Learning 2008.