Normative models of attention

Wei Ji Ma

New York University

Center for Neural Science, Department of Psychology, Center for Data Science, Institute for the Study of Decision Making

20171013

Deluges of data

David Teniers the Younger,
Archduke Leopold Wilhelm in his gallery in Brussels (around 1651)

- **Distributed attention**: the more items need to be attended to, the lower the precision with which each is encoded.
- **Selective attention**: for a given number of items, those with greater relevance are encoded with higher precision.
- **Attention for choice:** in deliberate choice between goods, attended goods are more likely to be chosen.

David Teniers the Younger,
Archduke Leopold Wilhelm in his gallery in Brussels (around 1651)

Models of distributed attention, selective attention, and attention for choice are mostly descriptive, even when quantitative.

In the psychology of attention, there is a lack of normative models.

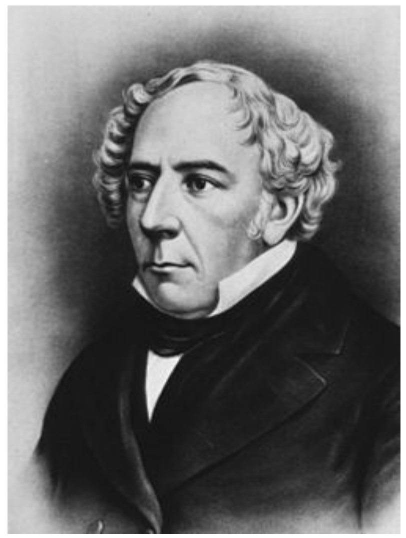
Three projects in progress

1. The effects of **set size** (number of items) on encoding precision: a new normative theory for existing data
2. The effects of **relevance** (priority) on encoding precision and confidence: new data and models, some normative
3. The effects of **fixation** (attention) on choice: a new normative theory for existing data

Part 1: The effects of **set size** (number of items) on encoding precision: a new normative theory

Ronald van den Berg
University of Uppsala

A very old observation



Sir William Hamilton
(1788-1856)

LECTURES

ON

METAPHYSICS

BY

SIR WILLIAM HAMILTON, BART.

PROFESSOR OF LOGIC AND METAPHYSICS IN THE UNIVERSITY OF EDINBURGH.

EDITED BY THE

REV. HENRY LONGUEVILLE MANSEL, B. D., OXFORD,

AND

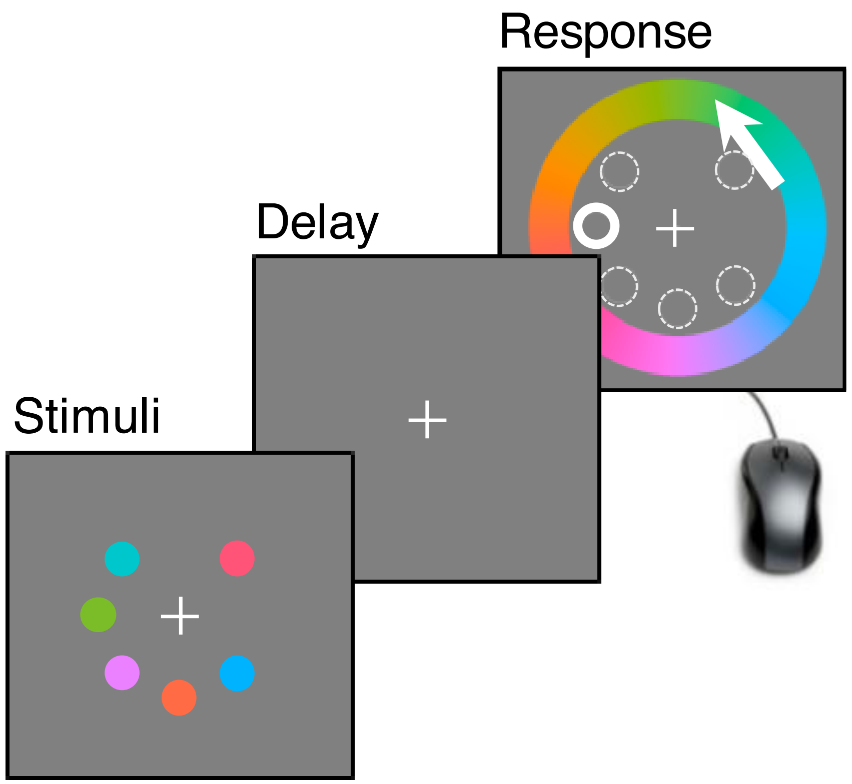
JOHN VEITCH, M. A., EDINBURGH.

BOSTON:
GOULD AND LINCOLN,
59 WASHINGTON STREET.
NEW YORK: SHELDON AND COMPANY.
CINCINNATI: GEORGE S BLANCHARD.
1859.

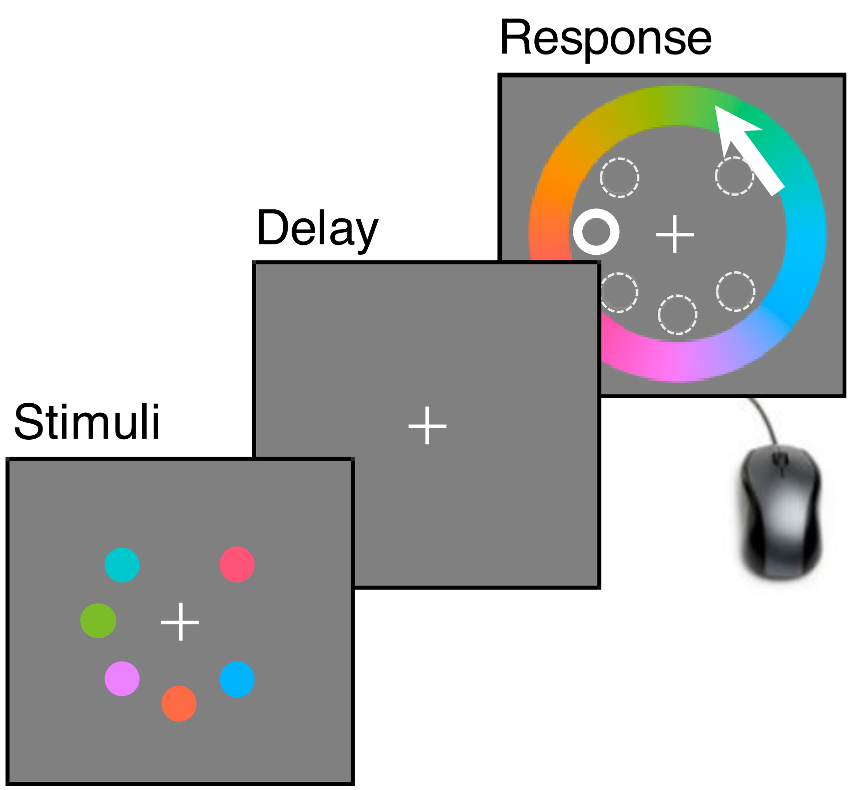
How many objects can the mind embrace at once?

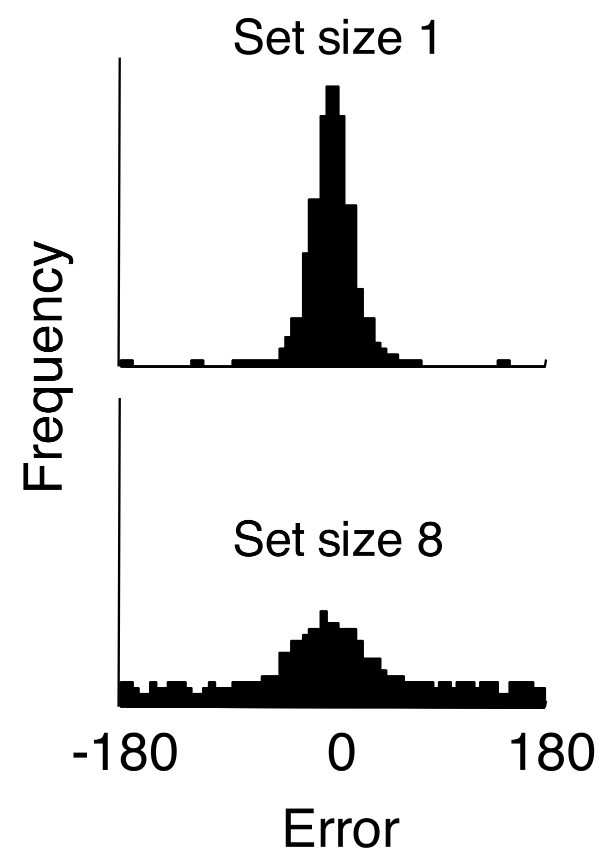
Supposing that the mind is not limited to the simultaneous consideration of a single object, a question arises, How many objects can it embrace at once? You will recollect that I formerly stated that the greater the number of objects among which the attention of the mind is distributed, the feebler and less distinct will be its cognizance of each.

"Pluribus intentus, minor est ad singula sensus."

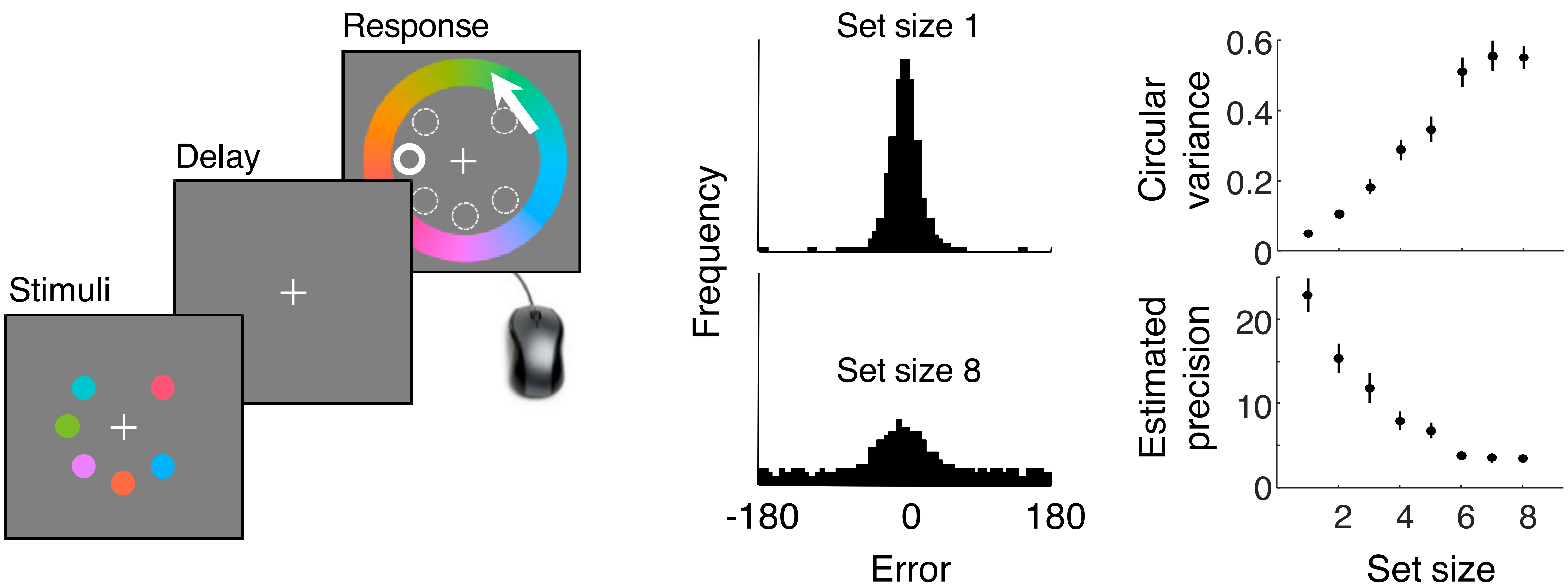

Response
Delay
Stimuli

Precision decreases with increasing set size





Precision decreases with increasing set size



But *why*?

Normative idea: two conflicting goals

“The need for energy management provides an interesting physiological perspective on a traditional view of attention as adaptation to the brain’s limited capacity to process information: energy limitations require that only a small fraction of the machinery can ever be engaged concurrently.”
— Lennie, *The cost of cortical computation,* 2003

“In this study we investigated the possibility that covert attention helps to control the expenditure of cortical computation by trading contrast sensitivity across attended and unattended areas of the visual field, even with impoverished displays and simple tasks.”
— Pestilli and Carrasco, 2005

Expected total loss = Expected behavioral loss + λ · Expected neural loss

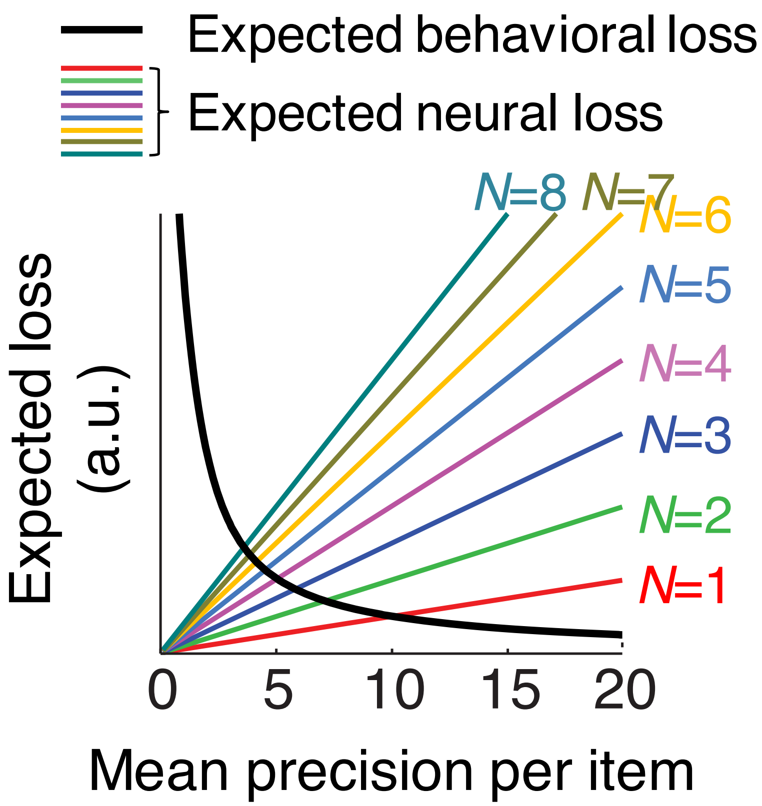
Implementation

- Stimulus $s \in [0,2\pi)$
- Noisy encoding: $x \sim \text{VonMises}(s,\kappa)$
- Precision $J = f(\kappa)$, f monotonic
- Variable precision: $J \sim \text{Gamma}\left(\frac{\bar{J}}{\tau},\tau\right)$
- Estimation error ε between x and s

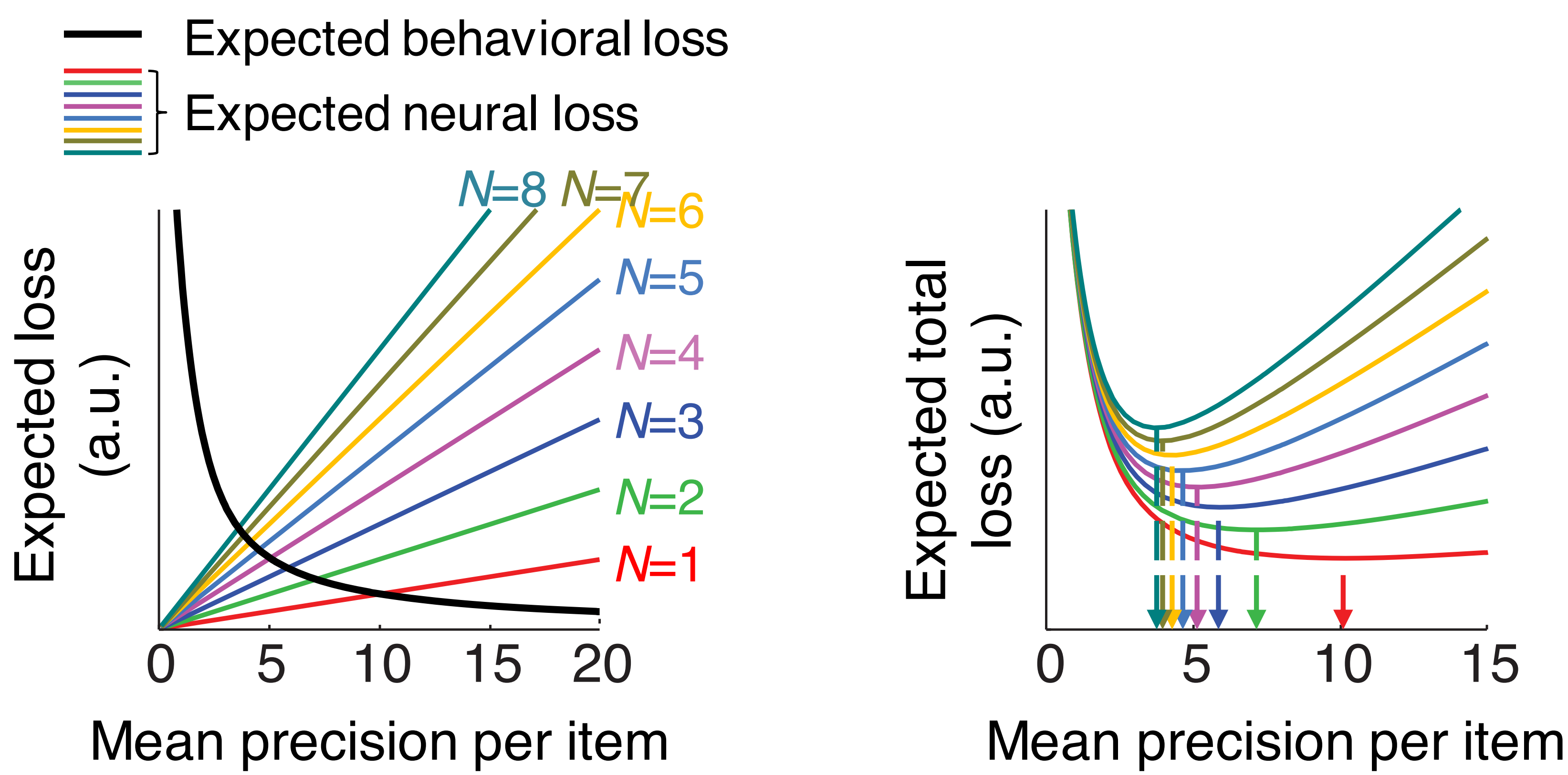
$$\text{Expected total loss} = \text{Expected behavioral loss} + \lambda \cdot \text{Expected neural loss}$$

$$= \left\langle \left|\varepsilon\right|^{\beta} \right\rangle_{p\left(\varepsilon|\bar{J}\right)} + \tilde{\lambda} N \bar{J}$$

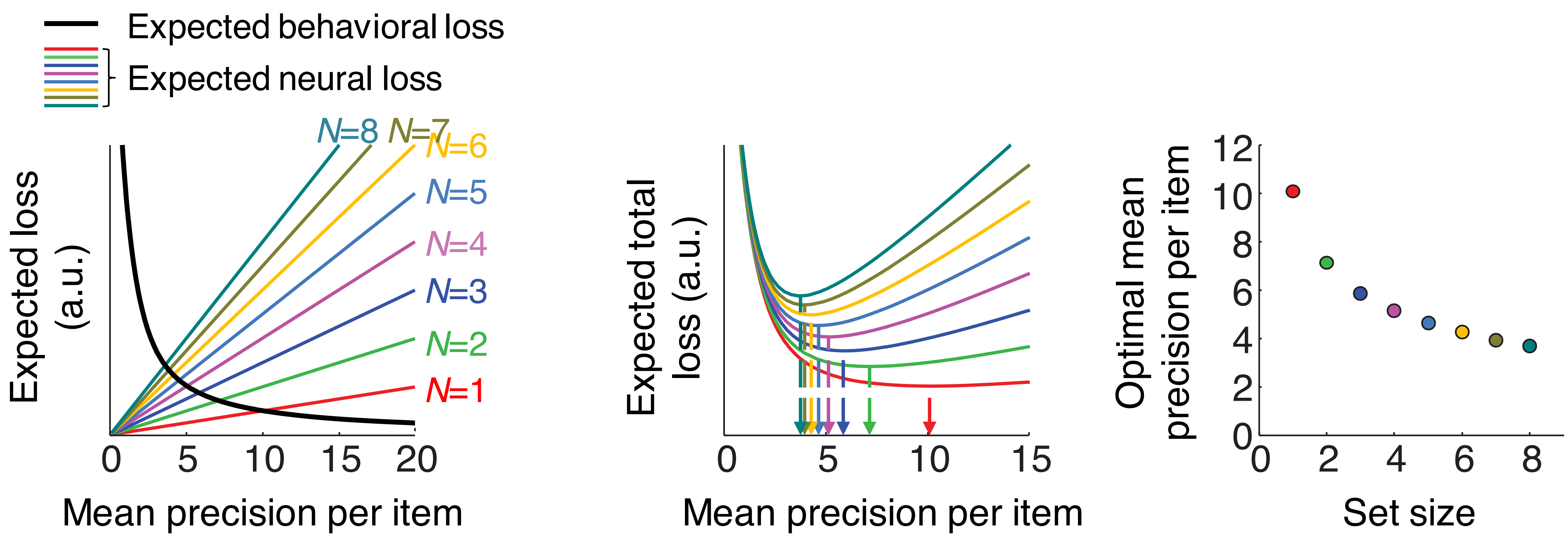
Model has only three parameters.



Expected behavioral loss
Expected neural loss
N=8
N=7
N=6
N=5
N=4
N=3
N=2
N=1
Expected loss (a.u.)
0
5
10
15
20
Mean precision per item



Expected behavioral loss
Expected neural loss
N=8
N=7
N=6
N=5
N=4
N=3
N=2
N=1
Expected loss (a.u.)
0
5
10
15
20
Mean precision per item
Expected total loss (a.u.)
0
5
10
15
Mean precision per item



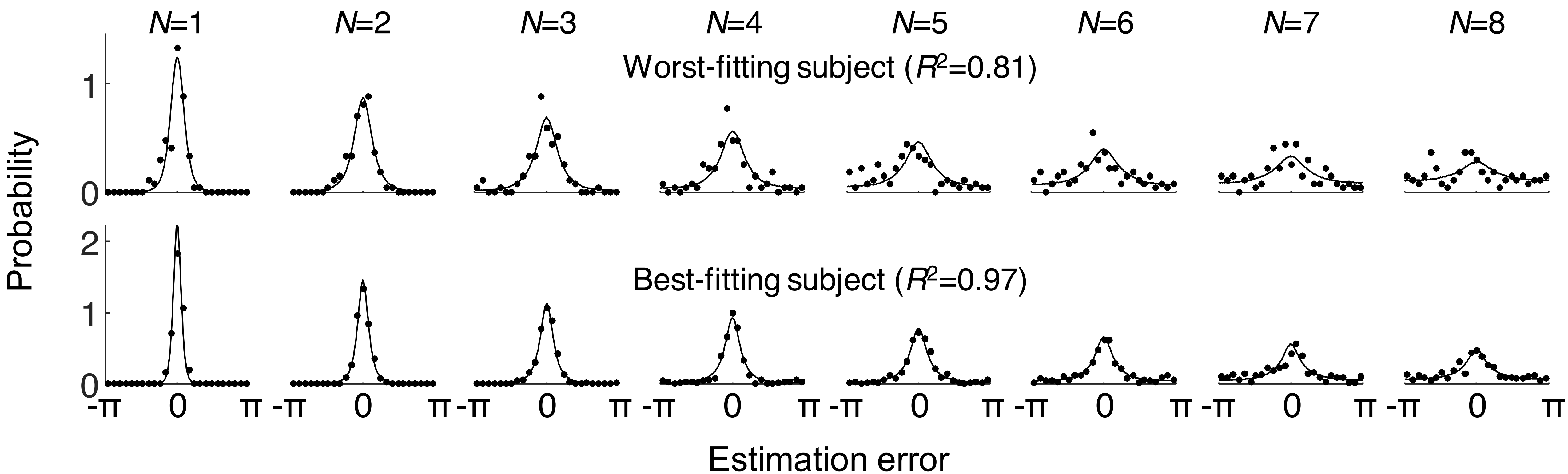
In general not a power law!

Experiment	Reference	Task	Feature	Set sizes	#subj
DE1	(Wilken & Ma, 2004)	Delayed estimation	Color	1, 2, 4, 8	15
DE2	(Zhang & Luck, 2008)	Delayed estimation	Color	1, 2, 3, 6	8
DE3	(Bays et al., 2009)	Delayed estimation	Color	1, 2, 4, 6	12
DE4	(van den Berg et al., 2012)	Delayed estimation	Orientation	1-8	6
DE5	(van den Berg et al., 2012)	Delayed estimation	Color	1-8	13
DE6	(van den Berg et al., 2012)	Delayed estimation	Color	1-8	13
CD1	(Keshvari et al., 2013)	Change detection	Color	1, 2, 4, 8	7
CD2	(Keshvari et al., 2013)	Change detection	Orientation	2, 4, 6, 8	10
CL1	(van den Berg et al., 2012)	Change localization	Color	2, 4, 6, 8	7
CL2	(van den Berg et al., 2012)	Change localization	Orientation	2, 4, 6, 8	11
VS	(Mazyar et al., 2013)	Visual search	Orientation	1, 2, 4, 8	6

Total 56,775 trials from 67 subjects

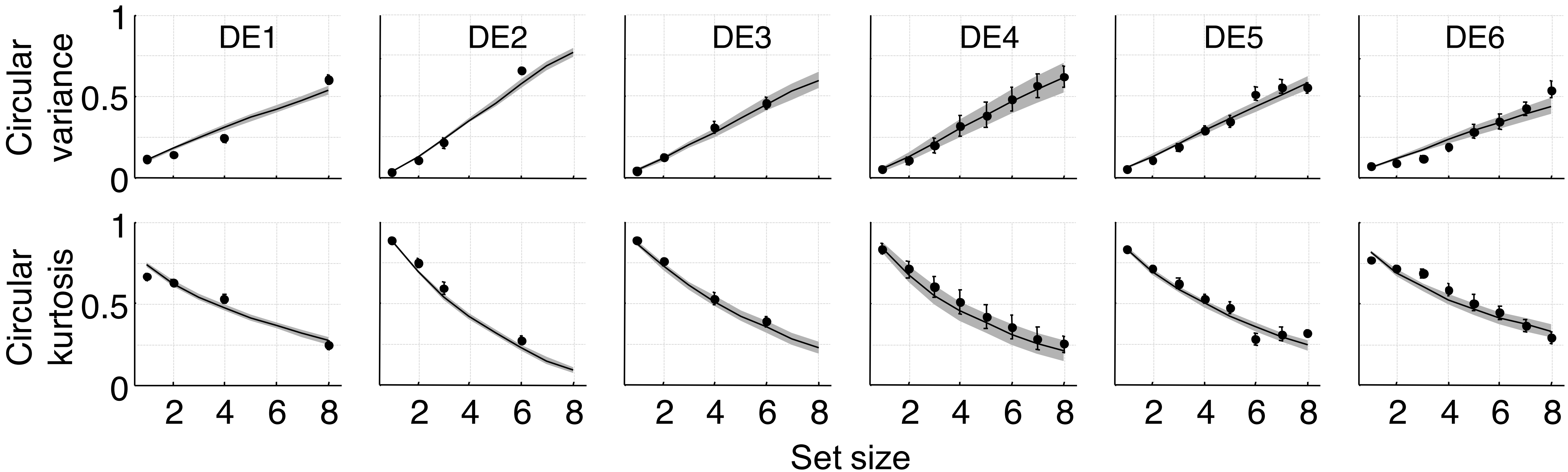
Van den Berg, Awh, and Ma, Psych Rev 2014

Model fits



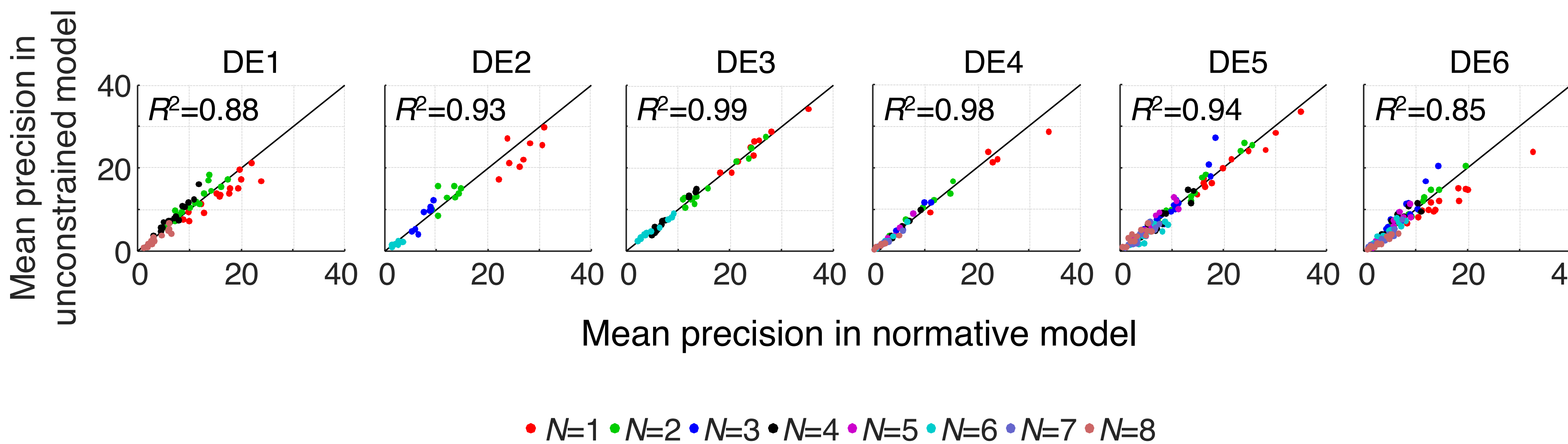
N=1
N=2
N=3
N=4
N=5
N=6
N=7
N=8
Worst-fitting subject (R^2=0.81)
Best-fitting subject (R^2=0.97)
Probability
1
0
2
1
0
-π
0
π
Estimation error

Model fits

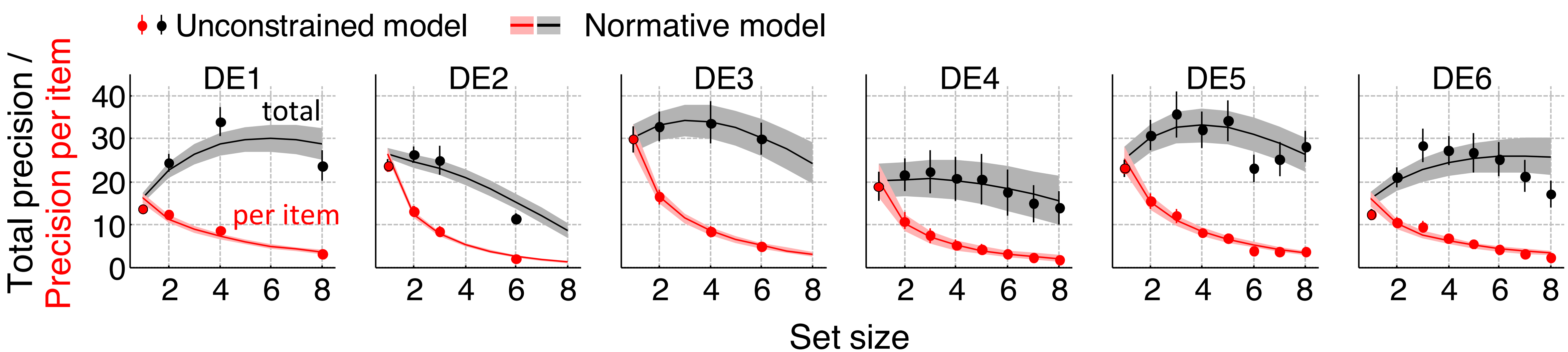


DE1
DE2
DE3
DE4
DE5
DE6
Circular variance
Circular kurtosis
0
0.5
1
2
4
6
8
Set size

Comparison with a maximally flexible model



Non-monotonicity of total precision



Conclusions Part 1

- The decrease of precision with set size in attention and working memory is usually thought of as a **cognitive limitation.**
- Instead, it might result from an **optimal trade-off** between behavioral and neural loss.
- Monetary incentives should increase attentional/working memory precision in a specific manner.
    - Surprising if you come from classic working memory.

Part 2: The effects of **relevance** (priority) on encoding precision and confidence: new data and models (some normative)

Aspen Yoo

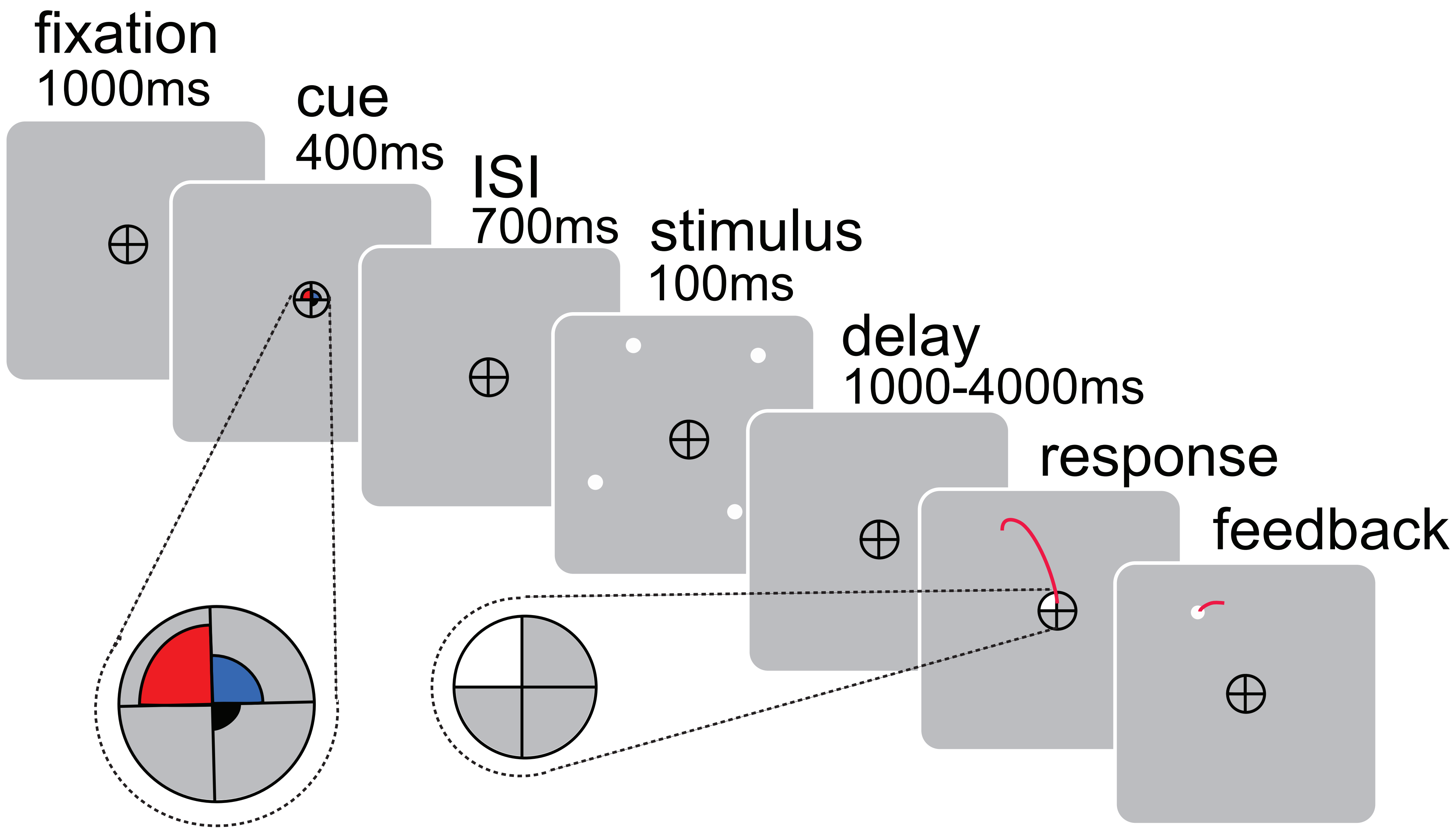
Zuzanna Klyszejko

Clay Curtis

Different items have different relevance

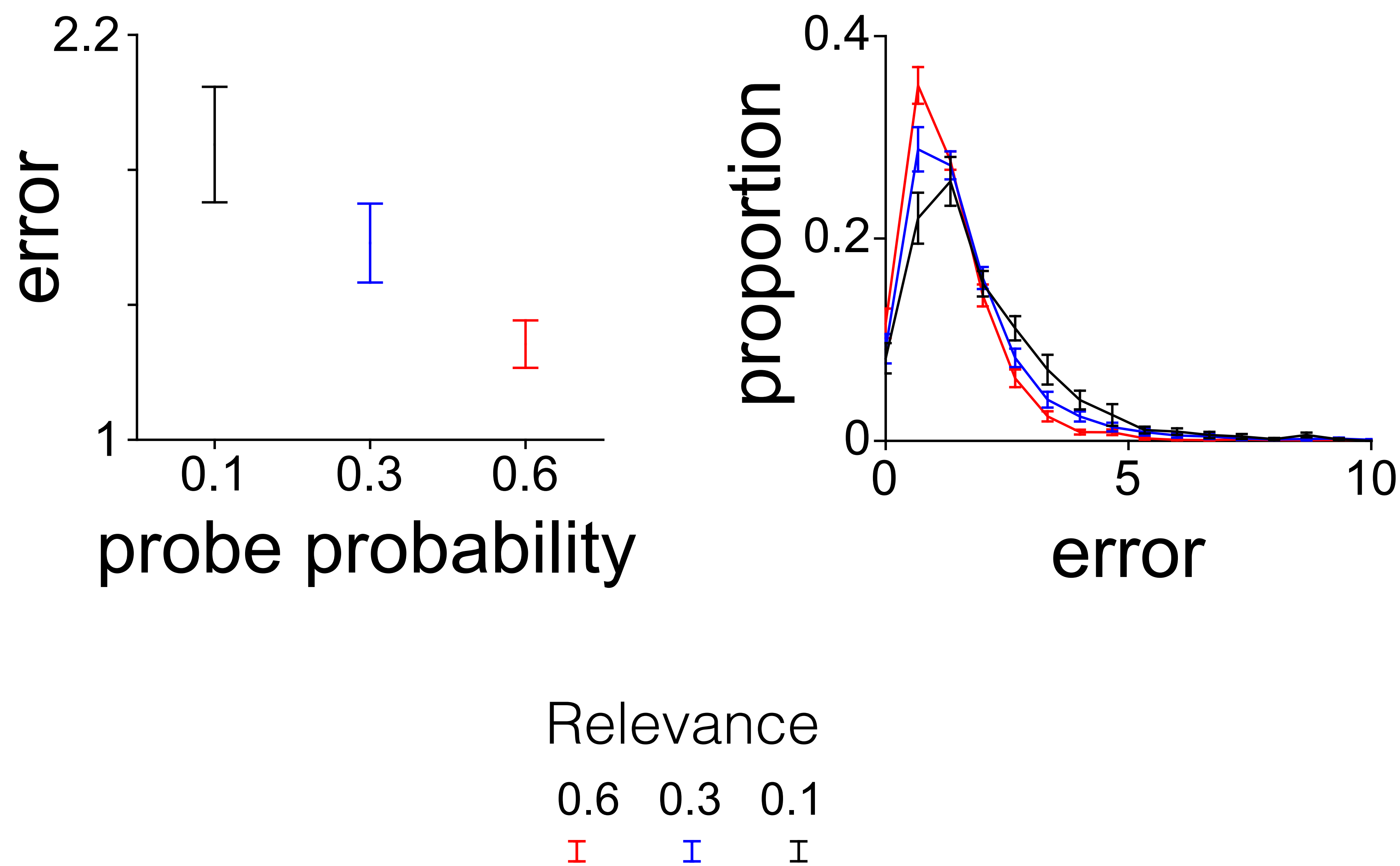
- Behavioral and neural evidence for allocation of resources based on relevance.
- Does attentional allocation optimally adjust to differences in relevance?

Experiment 1



Probe probability is **0.6**, **0.3**, or **0.1**.

Effects of the relevance manipulation



Models for attentional allocation

$$\text{Expected total loss} = \text{Expected behavioral loss}\left(\bar{J}_1, \bar{J}_2, \bar{J}_3\right) + \tilde{\lambda} N \bar{J}_{\text{total}}$$

now constant!

Mean precision allocated to the i^{th} item

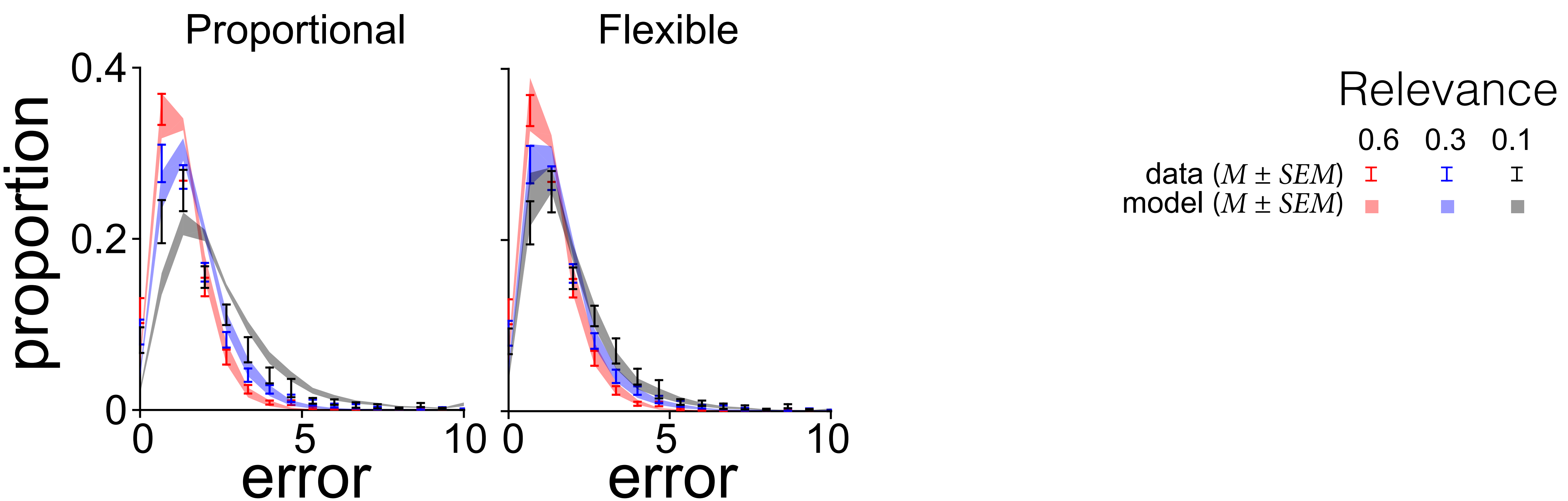
$$\bar{J}_i = p_i \bar{J}_{\text{total}}$$

- **Proportional model:** allocates attention in proportion to the probe probability (2 pars)
- **Flexible model:** proportions allocated can be anything (4 pars)
- **Normative model:** allocates attention to minimize expected behavioral loss (3 pars) - "rational inattention"

Model fits

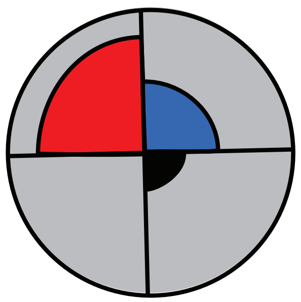

Proportional
0.4
0.2
0
proportion
0
5
10
error
Relevance
0.6
0.3
0.1
data ($M \pm SEM$)
model ($M \pm SEM$)

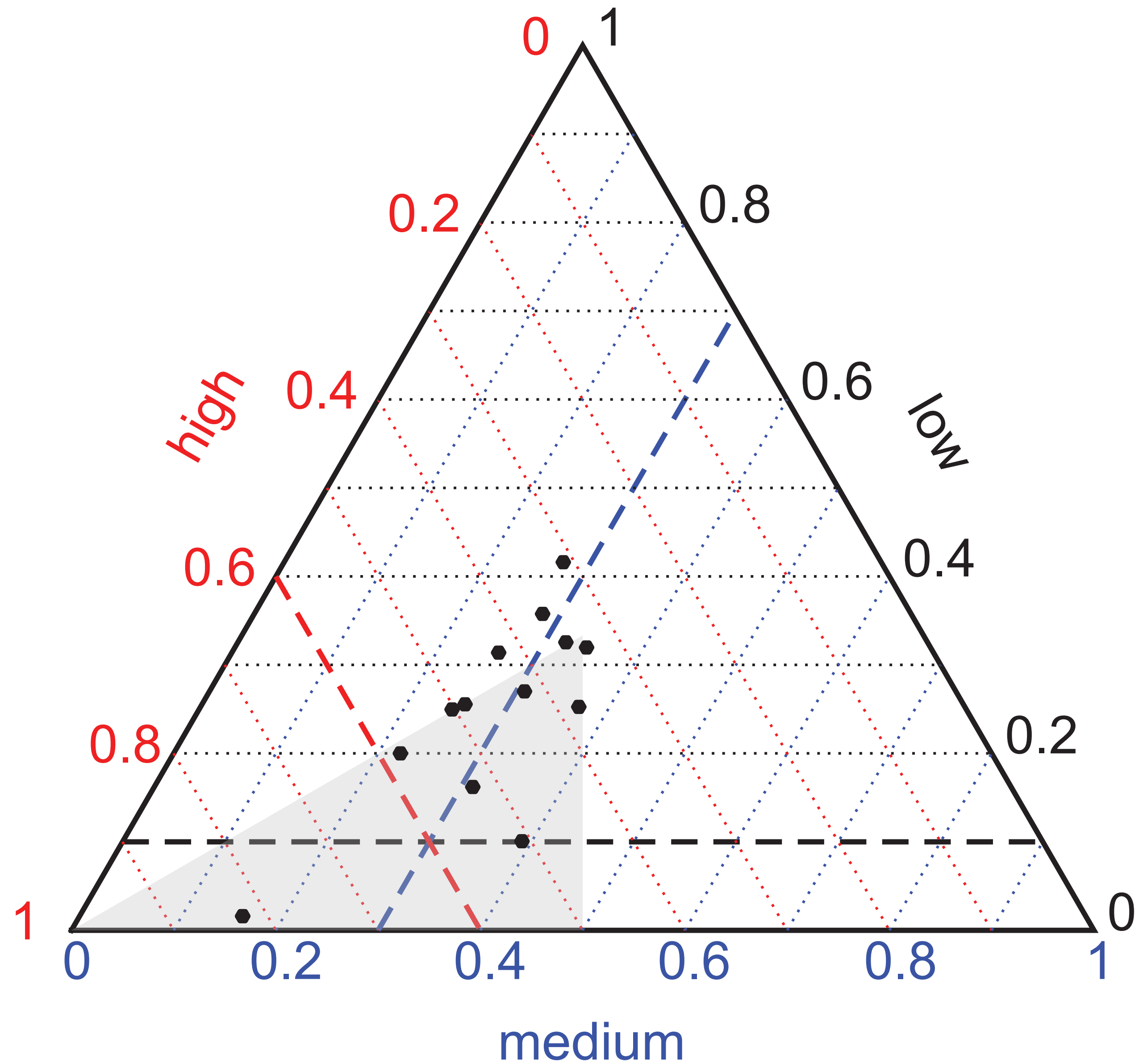
Model fits



Proportional
Flexible
proportion
0.4
0.2
0
0
5
10
0
5
10
error
error
Relevance
0.6
0.3
0.1
data ($M \pm SEM$)
model ($M \pm SEM$)

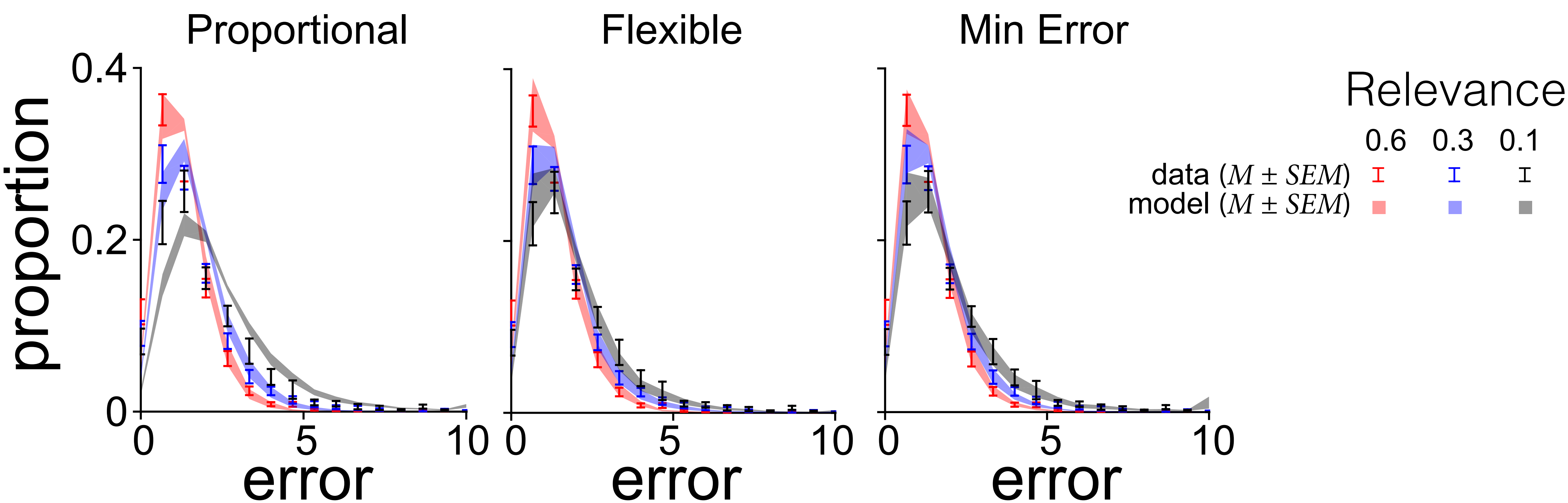
Attentional distribution in Flexible model



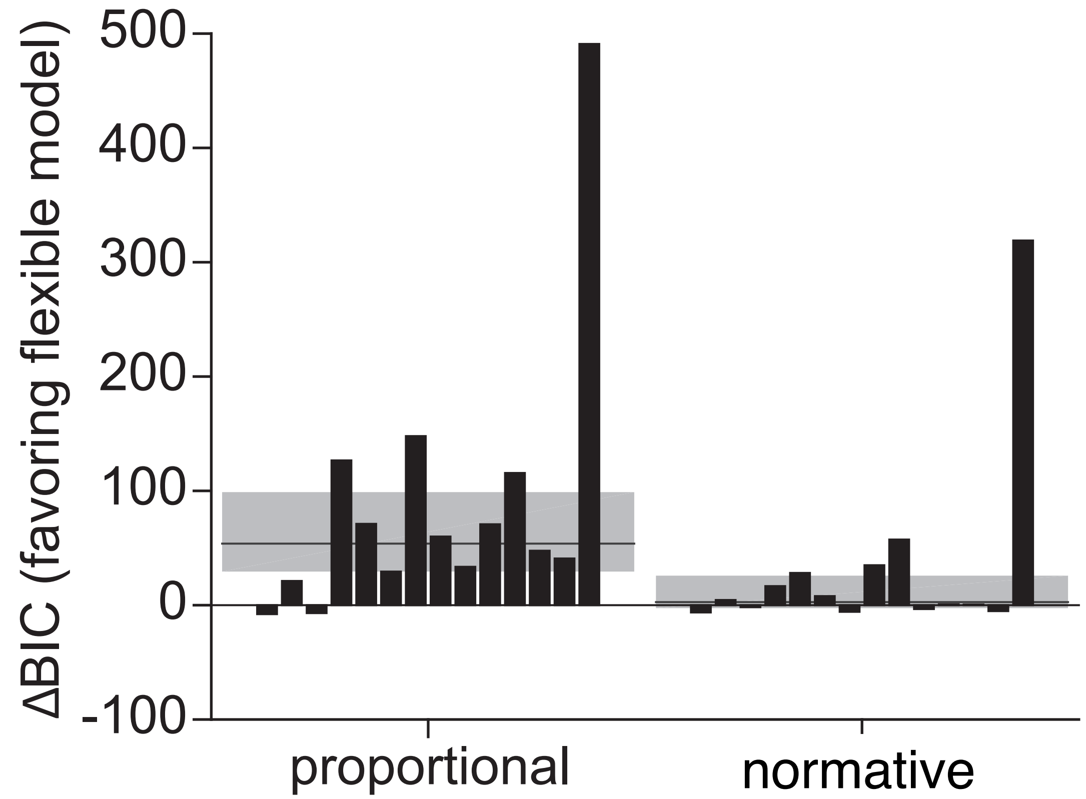


“Overallocation” to low, “underallocation” to high relevance

Model fits


Proportional
Flexible
Min Error
proportion
0.4
0.2
0
error
0
5
10
Relevance
0.6
0.3
0.1
data ($M \pm SEM$)
model ($M \pm SEM$)

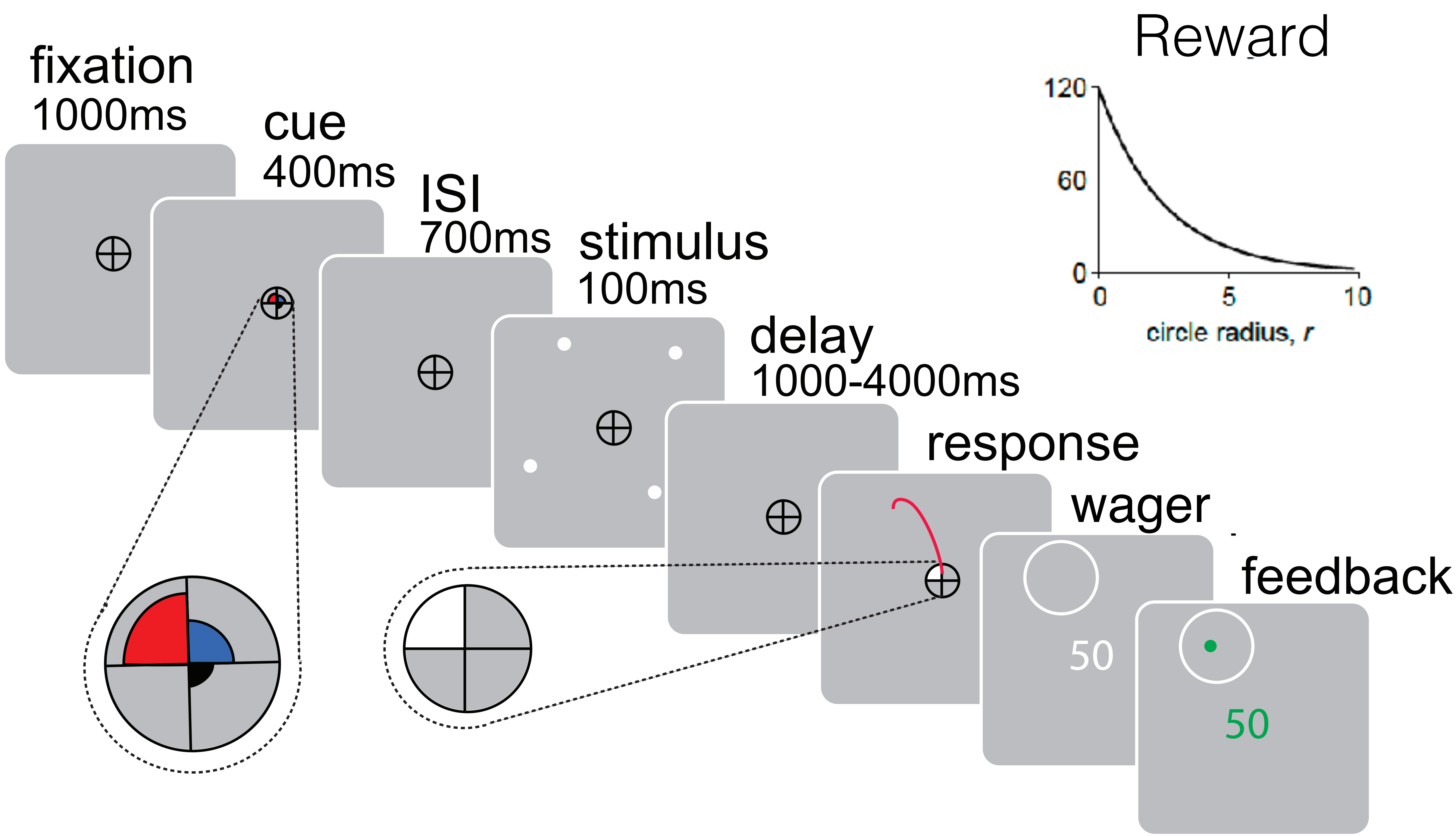
Model comparison



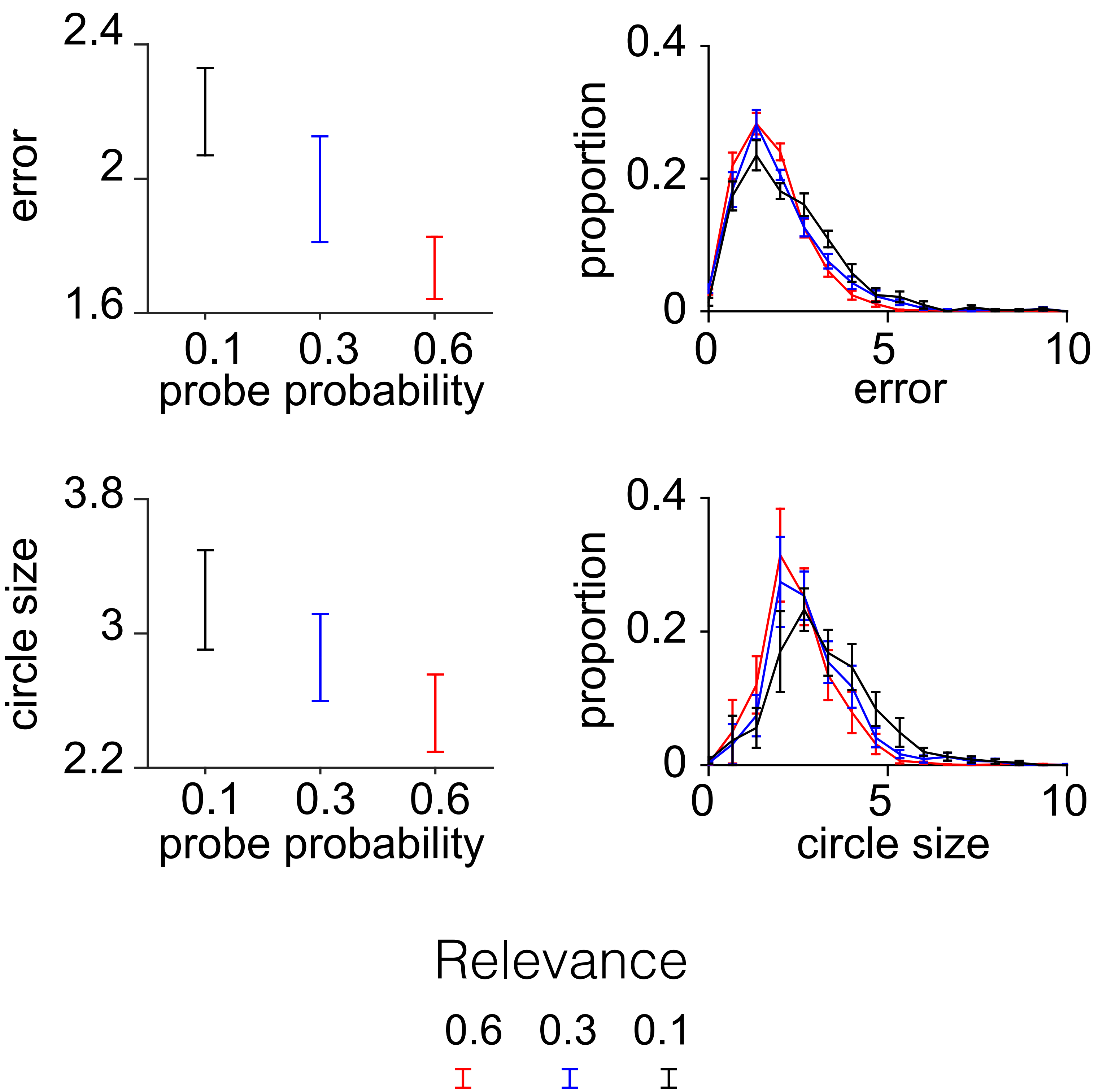
Attentional allocation minimizes error-based behavioral loss.

Extension: effects of selective attention on confidence / metacognition

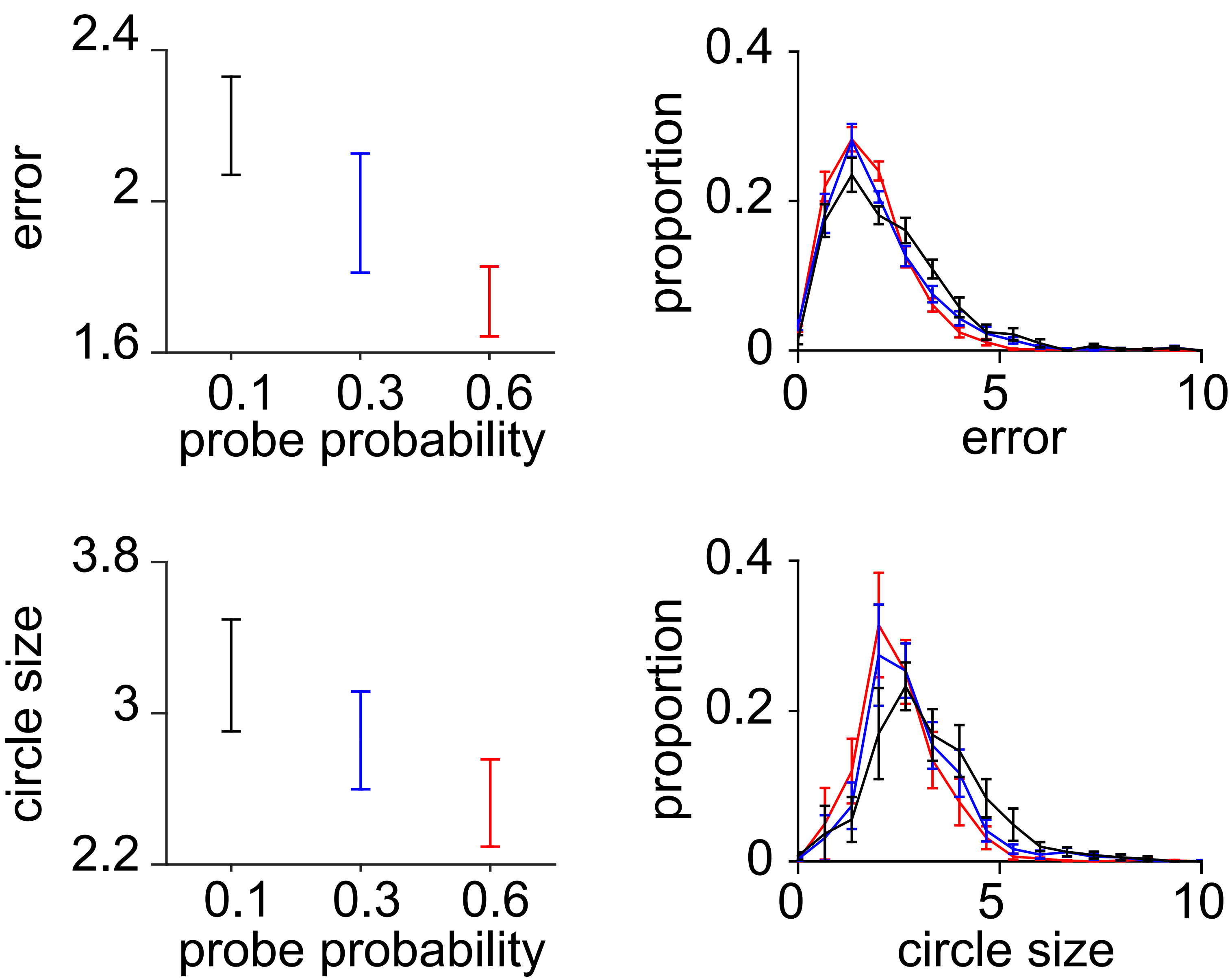
Experiment 2: post-decision wager



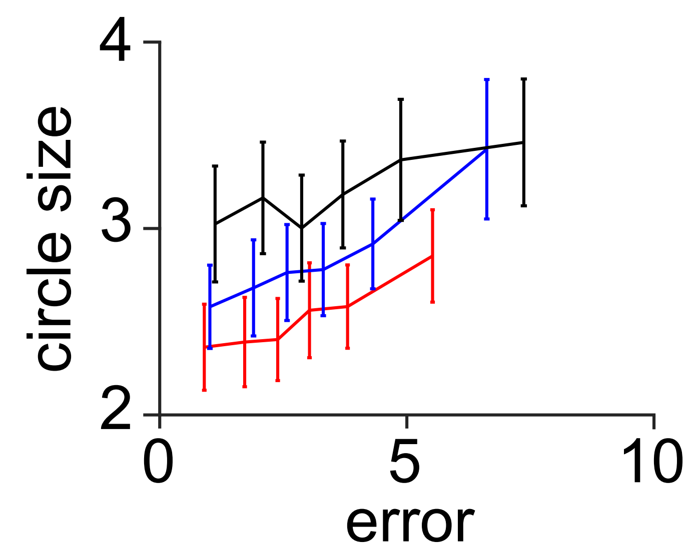
Effects of relevance



Effects of relevance



Internal variations in the quality of the representation!

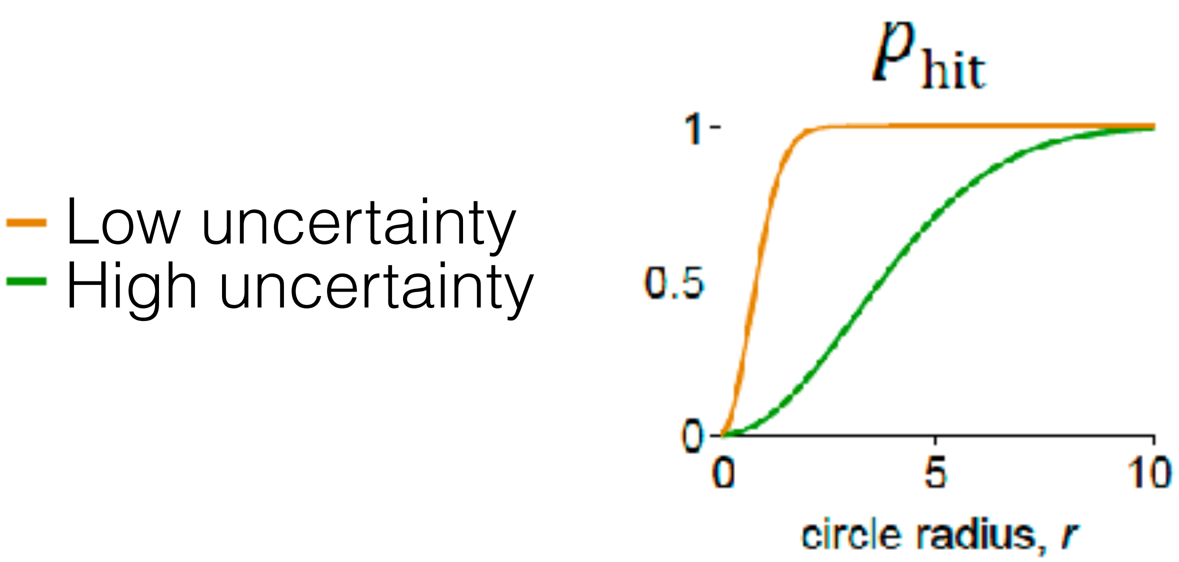


Relevance

0.6 0.3 0.1

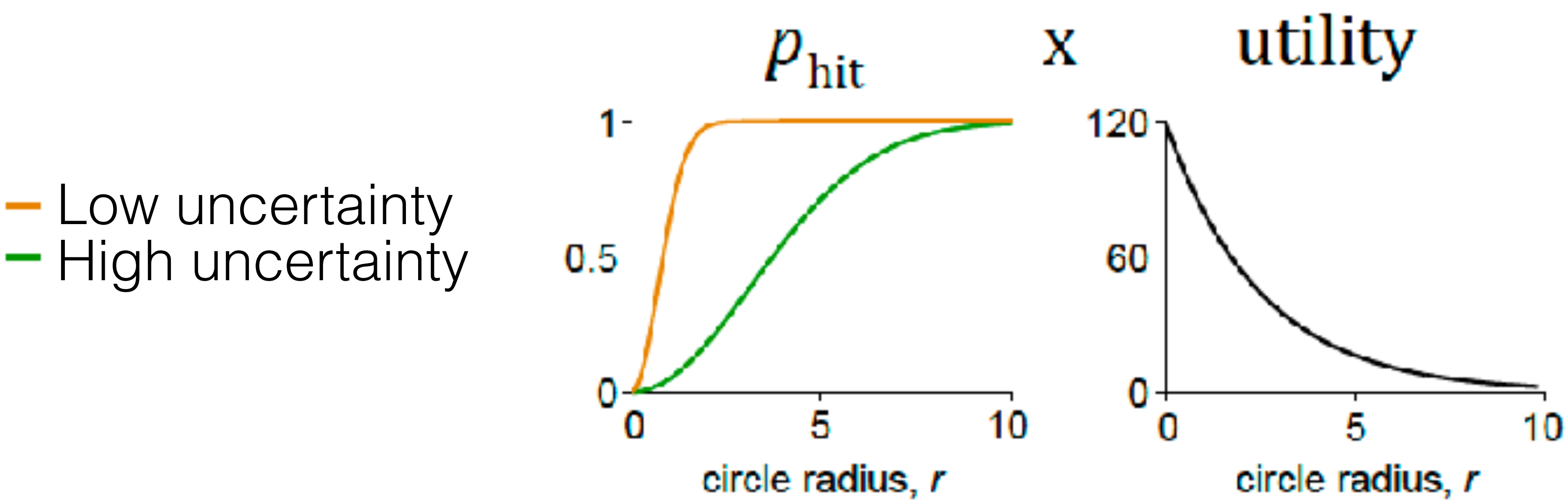
Model for post-decision wager

How to set the circle radius for given uncertainty?



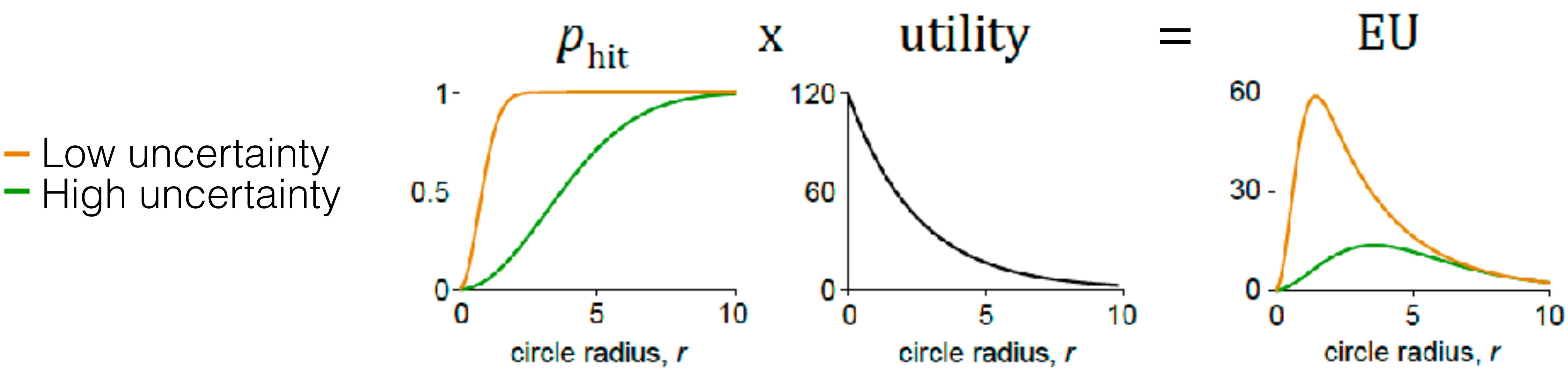
Model for post-decision wager

How to set the circle radius for given uncertainty?



Model for post-decision wager

How to set the circle radius for given uncertainty?



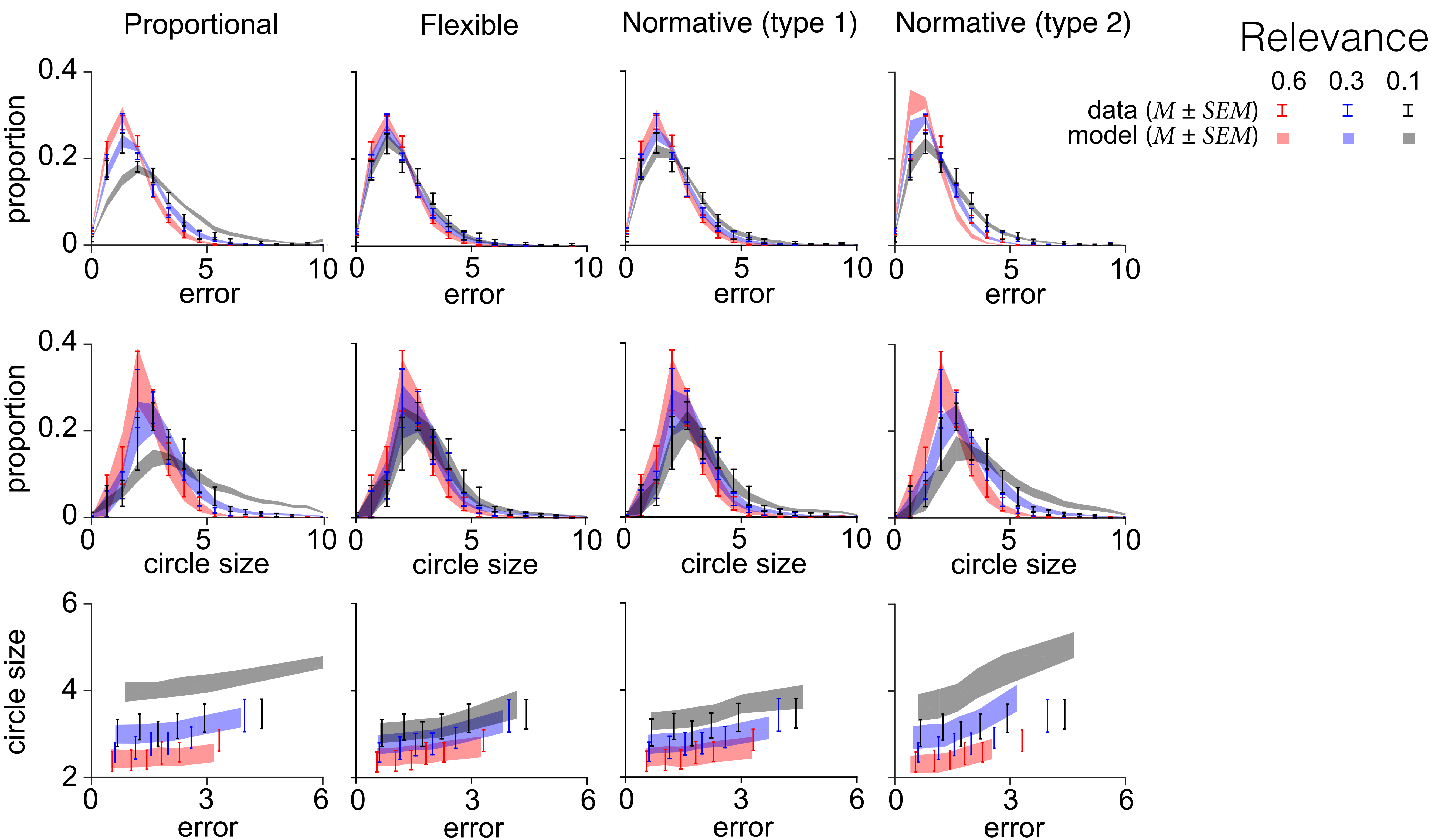
Reported circle radius: softmax read-out of EU

Requires access to trial-to-trial representation of uncertainty!

Models for attentional allocation (estimate and confidence)

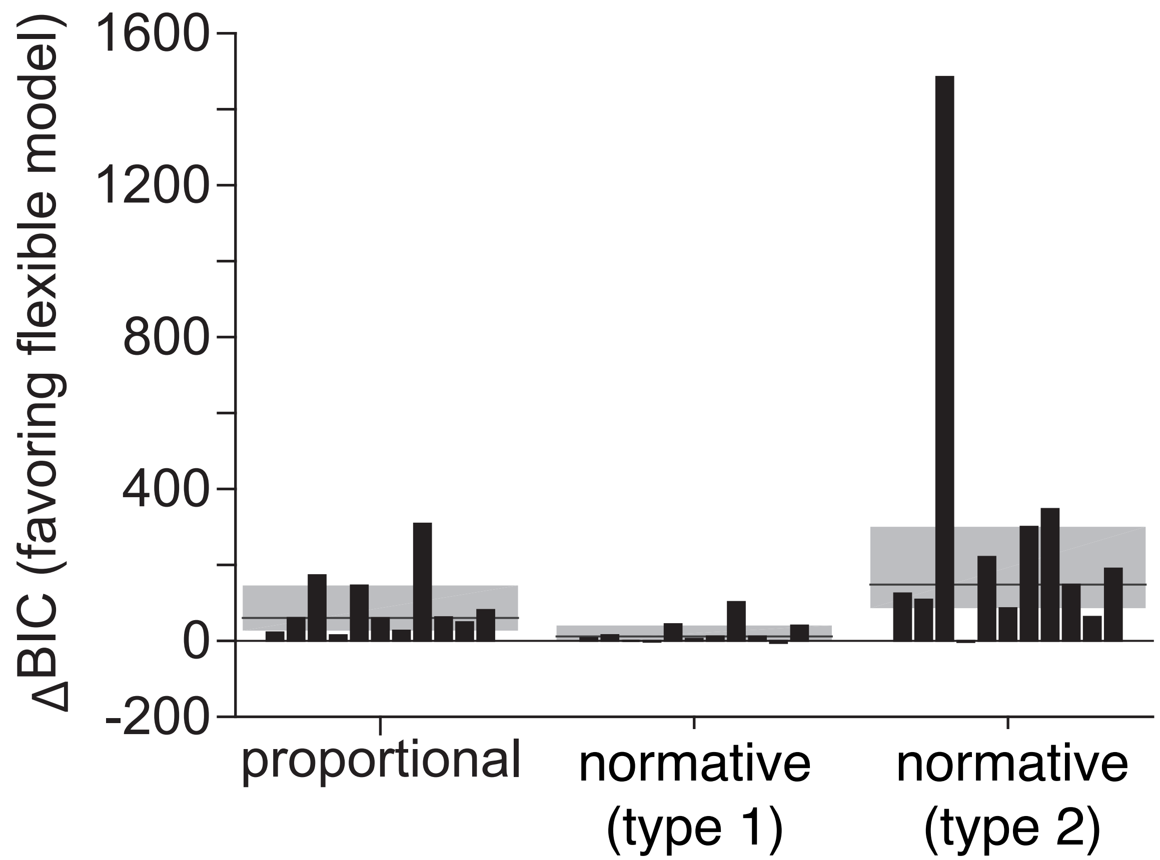
- **Proportional model:** allocates attention in proportion to the probe probability (4 pars)
- **Flexible model:** proportions allocated can be anything (6 pars)
- **Normative model (type 1):** allocates attention to minimize expected behavioral loss, wager is an afterthought (5 pars)
- **Normative model (type 2):** allocates attention to maximize expected point gain from wager (4 pars)

Model fits



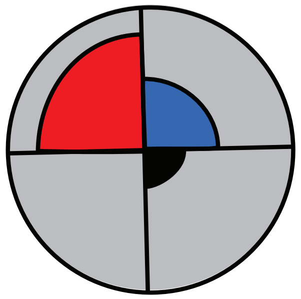
Proportional
Flexible
Normative (type 1)
Normative (type 2)
Relevance
0.6 0.3 0.1
data ($M \pm SEM$)
model ($M \pm SEM$)
proportion
error
circle size
0.4
0.2
0
5
10
6
4
2
3

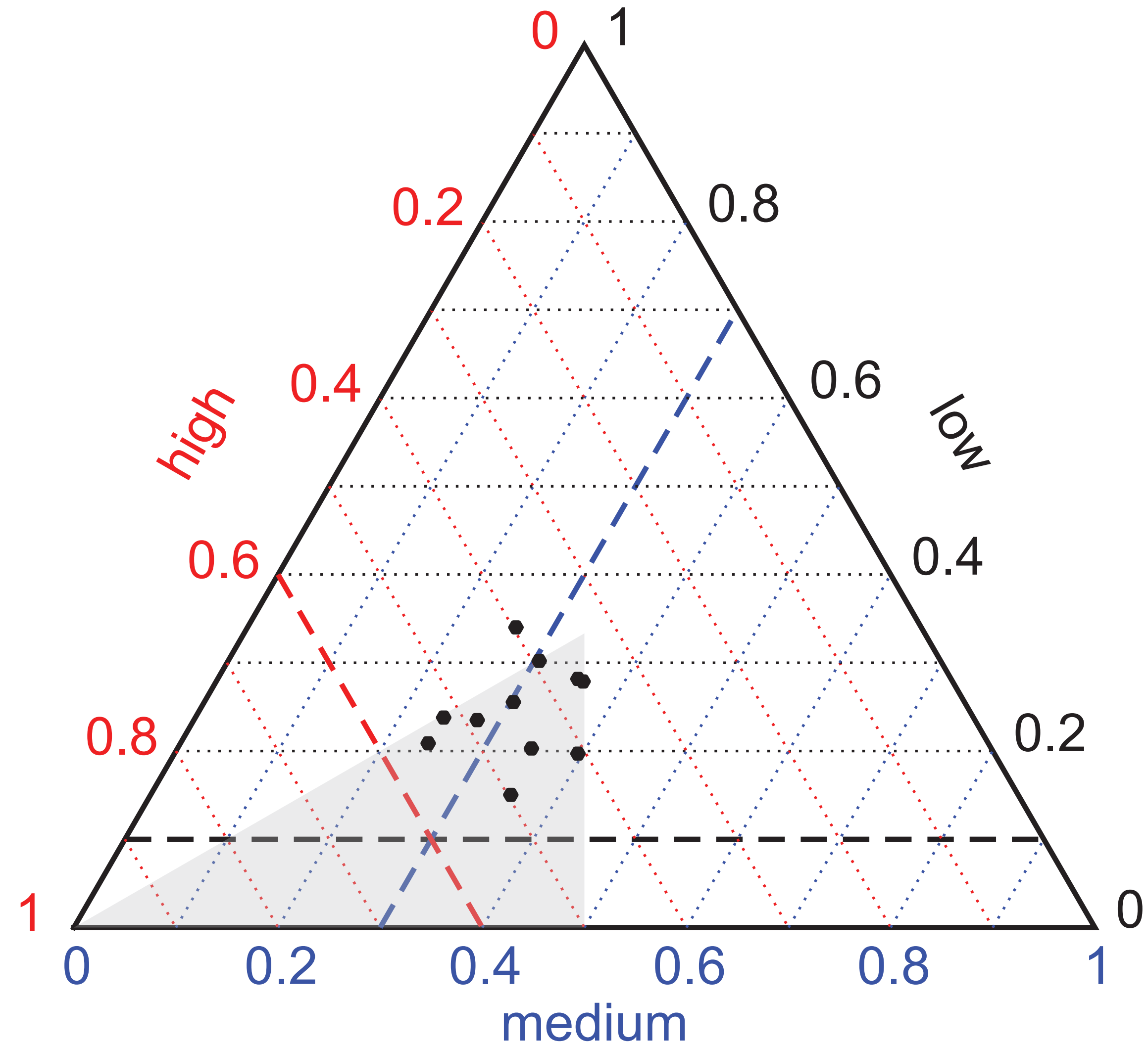
Model comparison



Attentional allocation minimizes expected error-based loss, and not point gain from a post-decision wager.

Attentional distribution in the Flexible model





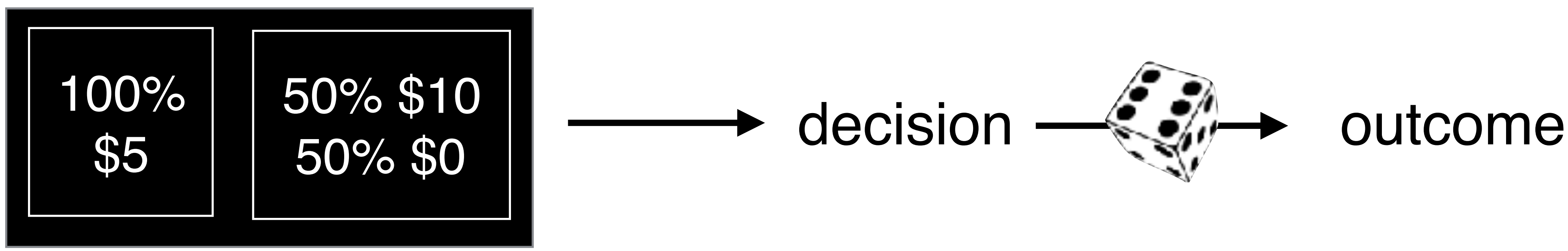
Conclusions Part 2

- Observers allocate more resources to items with higher relevance
  - **Not proportional**: overallocation to low, underallocation to high relevance
  - Apparently to **minimize error-based expected loss**.
- *Allocation strategy* did not change when a post-decision wager was introduced,
  - but *wager decision* itself might be maximizing expected utility.
- Working memory **stores uncertainty** on a trial-by-trial basis.

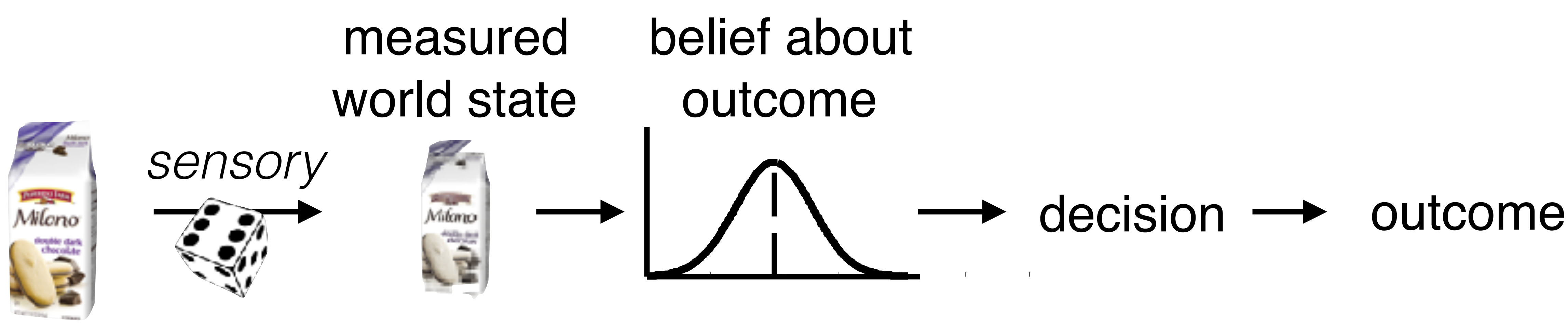
Part 3: The effects of **fixation** on choice: a new normative theory for Krajbich & Rangel data

Zhiwei Li

Uncertainty about outcome



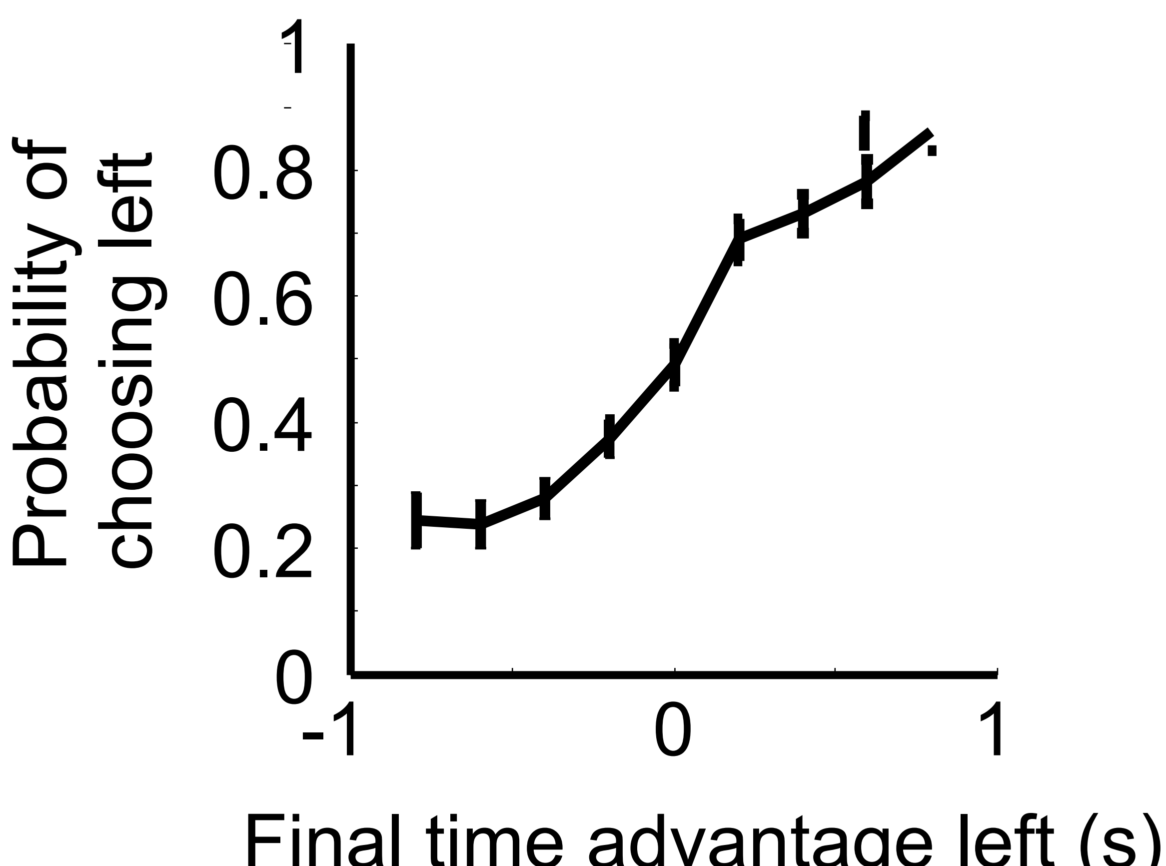
Uncertainty about outcome derived from uncertainty about world state



In value-based decision-making, do people build belief distributions over world states like in perceptual decision-making?

Krajbich, Armel, Rangel 2010

free viewing, eye tracked → decision



People more often chose the item they fixated on for longer.

Can an inference model explain this?

Value inference model (VIM)

Step 1: From observations to posterior

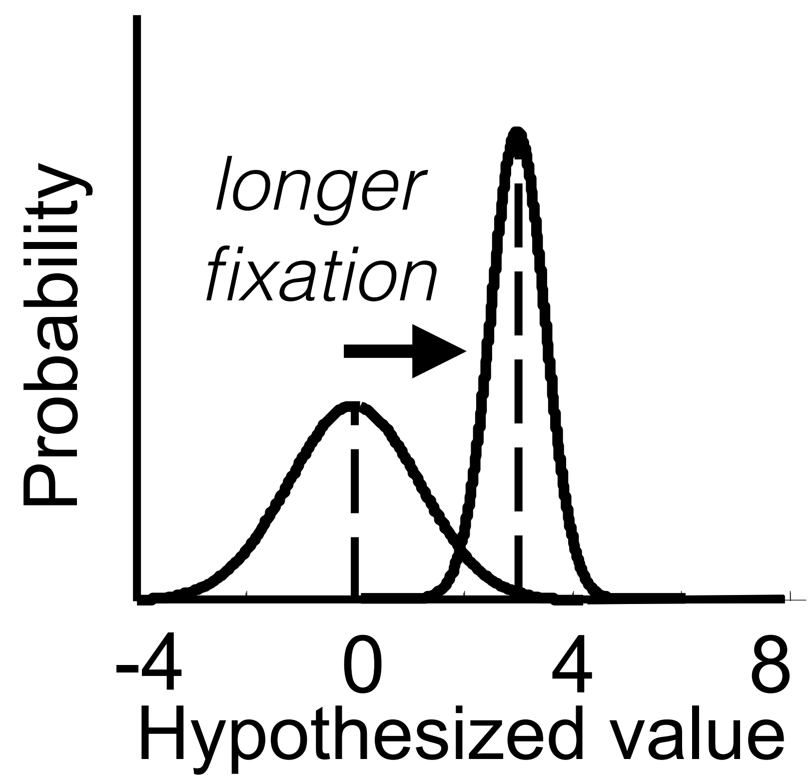
Step 2: From posterior to utility

Step 3: From utility to choice

VIM Step 1: From observations to posterior

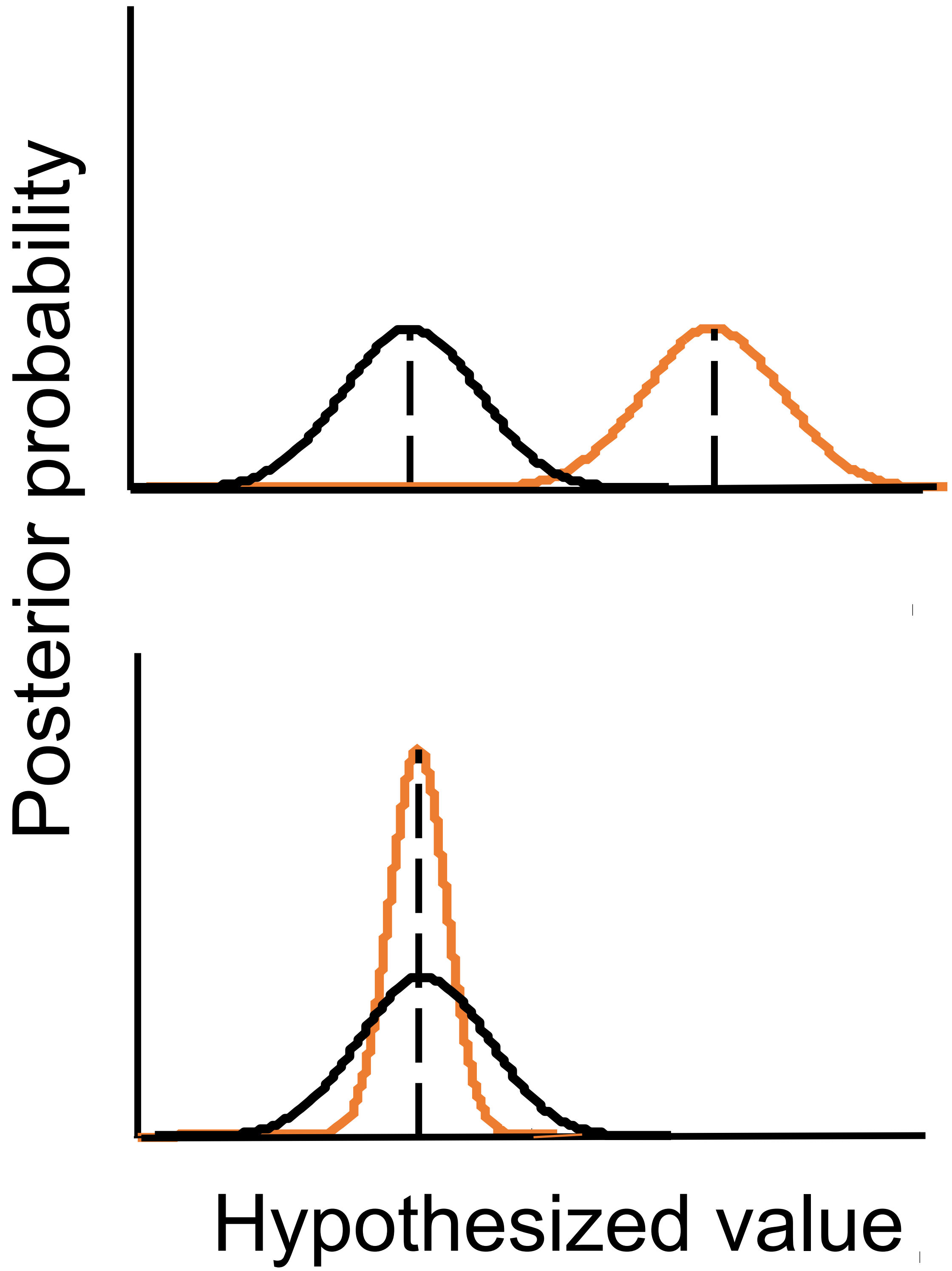
Uncertainty decreases over time

In part due to retrieval of memories (Shadlen and Shohamy 2016)



Internal belief distribution, not histogram of experienced outcomes!

VIM Step 2: From posterior to utility



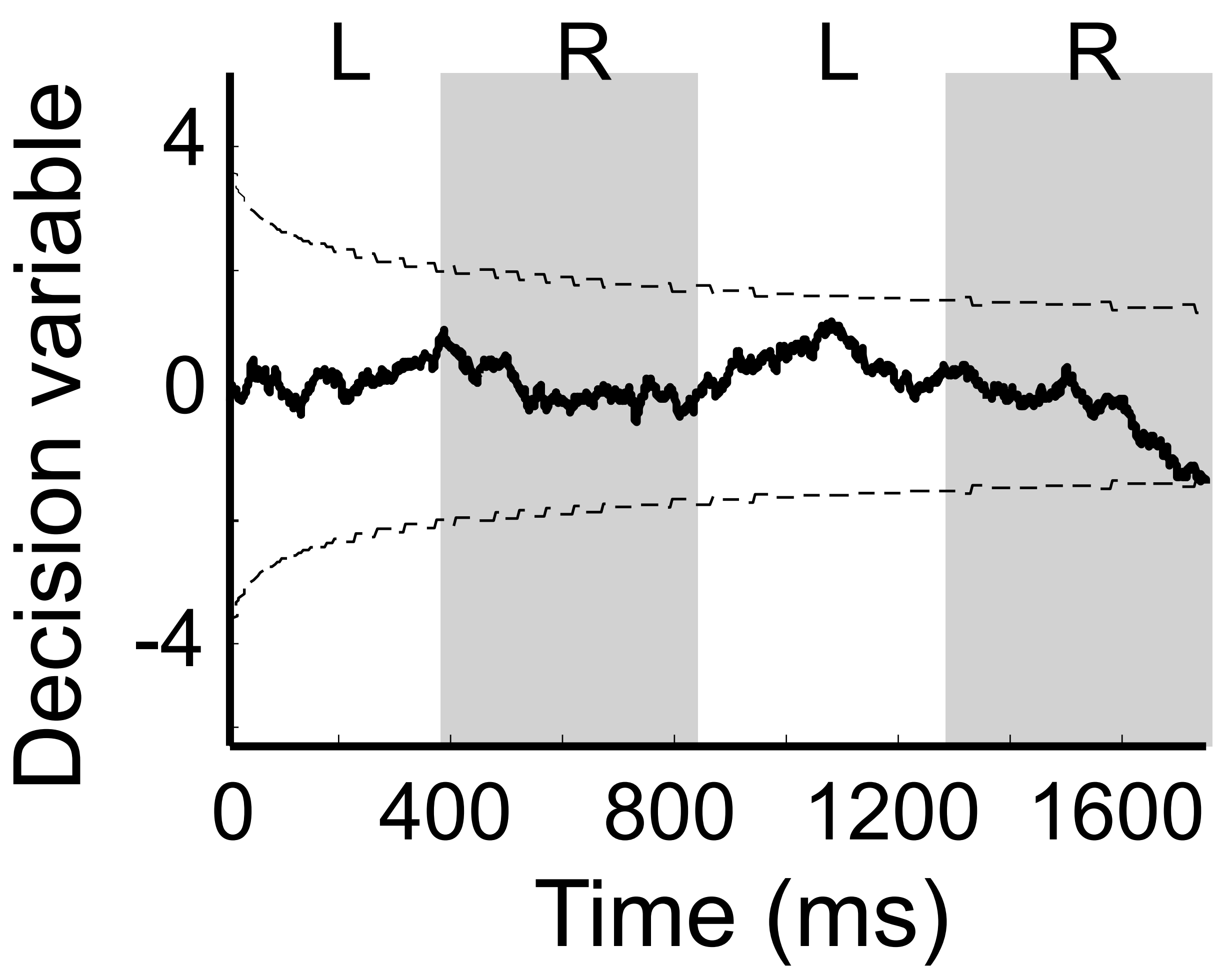
agent prefers higher posterior mean and lower uncertainty

$$U = \mu_{\text{posterior}} - A \cdot \sigma_{\text{posterior}}$$

VIM Step 3: From utility to choice

Decision variable: $DV = U(\text{left}) - U(\text{right})$

Decision is made when DV reaches a collapsing boundary



$$B_t = B_0 \cdot \exp\left(-\left(\frac{t}{\lambda}\right)^k\right)$$

Hawkins et al. 2015

Model comparison

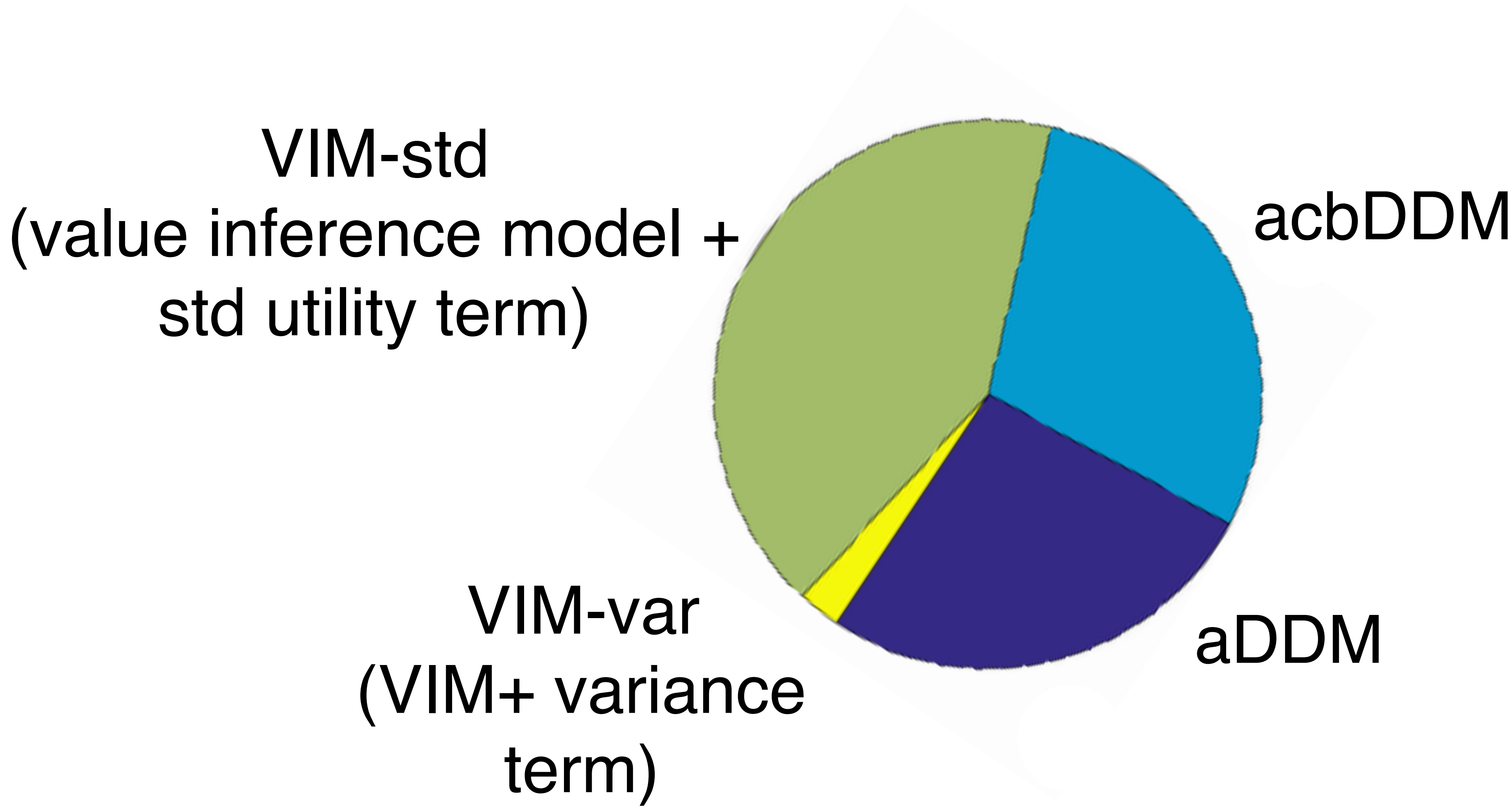
VIM: value inference model
aDDM: attention drift-diffusion model (Krajbich and Rangel)
acbDDM: aDDM with collapsing bound

	AICc	BIC
VIM-aDDM	**-903** (-1376, -473)	**-634** (-1125, -226)
acbDDM-aDDM	**-876** (-1318, -518)	**-607** (-1040, -237)
VIM-acbDDM	-27 (-198, 147)	-27 (-196, 144)

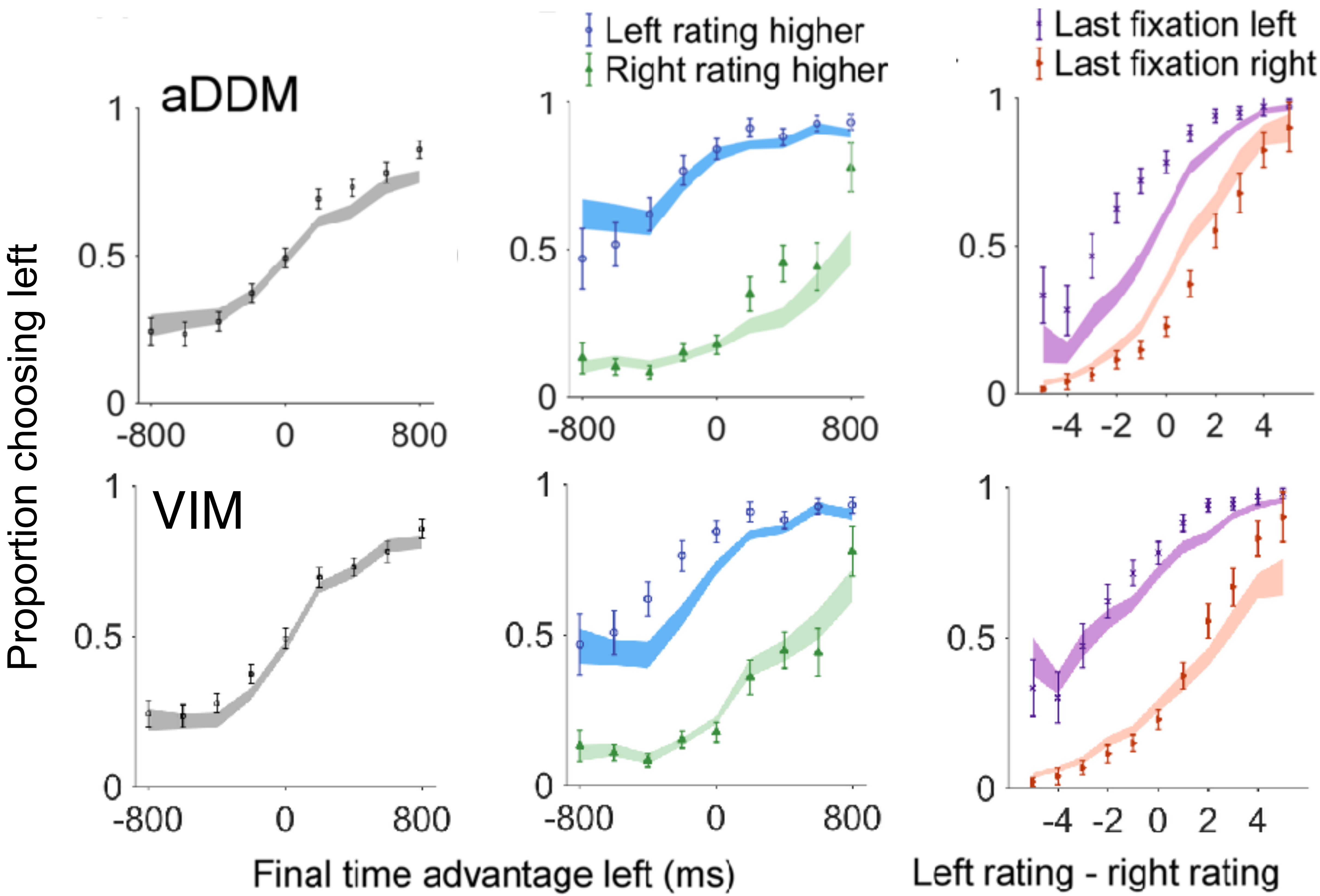
(median and bootstrapped 95% confidence interval)

Bayesian model selection for groups

Stephan et al. 2009; Rigoux et al. 2014



Model fits



aDDM
VIM
Left rating higher
Right rating higher
Last fixation left
Last fixation right
Proportion choosing left
Final time advantage left (ms)
Left rating - right rating
1
0.5
0
-800
0
800
-4
-2
0
2
4

Conclusion Part 3

- **Value inference** combined with **uncertainty aversion** might underlie fixation-based choice biases.
- Strong evidence for **collapsing bound** (regardless of VIM or aDDM)
- All models we tested show **systematic deviations** from the data.

Towards normative models of attention

- Distributed attention: minimize expected error-based loss while minimizing neural loss
- Selective attention: minimize expected error-based loss (neural loss fixed)
- Attention for choice: maximize posterior mean while minimizing posterior uncertainty

David Teniers the Younger,
Archduke Leopold Wilhelm in his gallery in Brussels (around 1651)