

*Testing the Bayesian Brain:*  
A Statistical Model for Sampling-Based  
Probability Estimates

---

Henrik Singmann

University of Zurich (from October: University of Warwick)

⒱ <http://singmann.org>

🐦 @HenrikSingmann



David Kellen  
Syracuse University

$$P(A|B) = \frac{P(A|B)P(A)}{P(B)}$$



Gamma



Normal

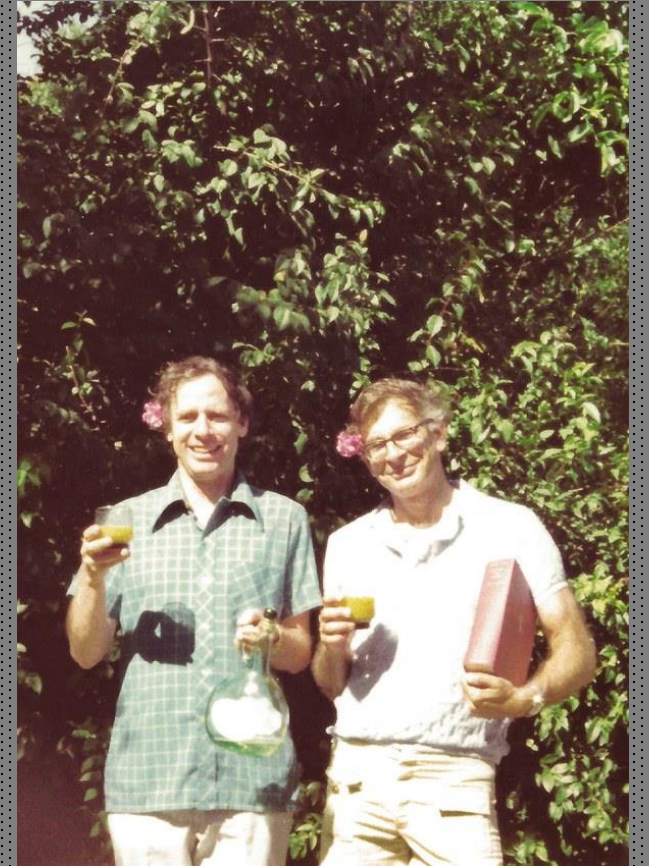


**Probabilities in the head**  
(Thomas Bayes)

**Judgment  
under  
uncertainty:  
Heuristics  
and biases**

Edited by  
DANIEL KAHNEMAN  
PAUL SLOVIC  
AMOS TVERSKY

People are  
not Bayesian!  
(at least in higher-  
level tasks)



# Resurgence of Bayesian Models since ca. 2000

Review

CellPress

## A Bayesian perspective on magnitude estimation

Frederike H. Petzschner<sup>1</sup>, Stefan Glasauer<sup>2,3,4</sup>, and Klaas E. Stephan<sup>1,5</sup>

<sup>1</sup>Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zürich & ETH Zürich, Switzerland

<sup>2</sup>Center for Sensorimotor Research and Department of Neurology, Ludwig-Maximilian University Munich, Munich, Germany

<sup>3</sup>German Center for Vertigo and Balance Disorders (DSGZ), Ludwig-Maximilian University Munich, Munich, Germany

<sup>4</sup>Bernstein Center for Computational Neuroscience, Ludwig-Maximilian University Munich, Munich, Germany

<sup>5</sup>Wellcome Trust Centre for Neuroimaging, University College London, London, UK

Our representation of the physical world requires judgments of magnitudes, such as loudness, distance, or time. Interestingly, magnitude estimates are often not veridical but subject to characteristic biases. These biases are strikingly similar across different sensory modalities, suggesting common processing mechanisms that are shared by different sensory systems. However, the search for universal neurobiological prin-

So far, however, attempts to model magnitude estimation have often led to modality-specific or effect-specific explanations [17]. By contrast, recently proposed Bayesian accounts of magnitude estimation have the potential to provide a more general explanation that covers a wide set of behavioral characteristics and transcends any specific modality [18–20]. This Bayesian framework suggests that behavioral phenomena of magnitude estimation, such as

Review

Cell  
PRESS

## Organizing probabilistic models of perception

Wei Ji Ma

Department of Neuroscience, Baylor College of Medicine, 1 Baylor Plaza, Houston TX 77030, USA

Probability has played a central role in models of perception for more than a century, but a look at probabilistic concepts in the literature raises many questions. Is being Bayesian the same as being optimal? Are recent Bayesian models fundamentally different from classic signal detection theory models? Do findings of near-optimal inference provide evidence that neurons compute with probability distributions? This review aims to disentangle these concepts and to classify empirical evidence accordingly.

### Decision-making in an uncertain world

In order to survive and thrive, all animals must derive knowledge about the world from sensory observations. A

above, over target presence, landing location, and life span. Since this knowledge is based on sensory observations, the probability distribution is a conditional distribution, which can be denoted by  $q(\text{world state} \mid \text{observations})$ .

Knowledge is not sufficient for organisms; actions are needed. The wildebeest might decide whether to stay put, the badminton player whether to attempt a return, and the actuary what premium to set. Cost or utility is associated with each combination of true world state and action, denoted by  $C(\text{world state}, \text{action})$ : if the badminton player does not attempt to return the shuttle, energy is saved, but at the cost of a point if the shuttle lands inside the court. For the observer, the expected cost of an action is a weighted

# Resurgence of Bayesian Models since ca. 2000

Review

CellPress

## A Bayesian perspective on magnitude estimation

Opinion

Cell  
PRESS

*Approaches to cognitive modeling*

## Probabilistic models of cognition: exploring representations and inductive biases

Thomas L. Griffiths<sup>1</sup>, Nick Chater<sup>2</sup>, Charles Kemp<sup>3</sup>, Amy Perfors<sup>4</sup> and Joshua B. Tenenbaum<sup>5</sup>

<sup>1</sup> Department of Psychology, University of California, Berkeley, 3210 Tolman Hall MC 1650, Berkeley CA 94720-1650, USA

<sup>2</sup> Division of Psychology and Language Sciences, University College London, Gower Street, London WC1E 6BT, UK

<sup>3</sup> Department of Psychology, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh PA 15213, USA

<sup>4</sup> School of Psychology, University of Adelaide, Level 4, Hughes Building, Adelaide, SA 5005, Australia

<sup>5</sup> Brain and Cognitive Sciences Department, Massachusetts Institute of Technology, Building 46-4015, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

Cognitive science aims to reverse-engineer the mind, and many of the engineering challenges the mind faces

with abstract principles that allow agents to solve problems posed by the world – the functions that minds per-

Psychological Review

© 2017 American Psychological Association  
0033-295X/17/\$12.00 http://dx.doi.org/10.1037/rev0000052

## Bayesian Models of Cognition Revisited: Setting Optimality Aside and Letting Data Drive Psychological Theory

Sean Tauber and Daniel J. Navarro  
University of New South Wales

Amy Perfors  
University of Adelaide

Mark Steyvers  
University of California, Irvine

Recent debates in the psychological literature have raised questions about the assumptions that underpin Bayesian models of cognition and what inferences they license about human cognition. In this paper we revisit this topic, arguing that there are 2 qualitatively different ways in which a Bayesian model could be constructed. The most common approach uses a Bayesian model as a normative standard upon which to license a claim about *optimality*. In the alternative approach, a *descriptive* Bayesian model need not correspond to any claim that the underlying cognition is optimal or rational, and is used solely as a tool for instantiating a substantive psychological theory. We present 3 case studies in which these 2 perspectives lead to different computational models and license different conclusions about human cognition. We demonstrate how the descriptive Bayesian approach can be used to answer different sorts of questions than the optimal approach, especially when combined with principled tools for model evaluation and model selection. More generally we argue for the importance of making a clear distinction between the 2 perspectives. Considerable confusion results when descriptive models and optimal models are conflated, and if Bayesians are to avoid contributing to this confusion it is important to avoid making normative claims when none are intended.

**Keywords:** Bayesian cognitive models, rational models, inductive reasoning, generalization, optimal

inference provide evidence that neurons combine probability distributions? This review aims to clarify these concepts and to classify empirical findings accordingly.

making in an uncertain world

to survive and thrive, all animals must derive inferences about the world from sensory observations. A

the badminton player whether to attempt a return, and the actuary what premium to set. Cost or utility is associated with each combination of true world state and action, denoted by  $C(\text{world state}, \text{action})$ : if the badminton player does not attempt to return the shuttle, energy is saved, but at the cost of a point if the shuttle lands inside the court. For the observer, the expected cost of an action is a weighted



# Resurgence of Bayesian Models since ca. 2000

Trends in Cognitive Sciences

## Opinion

# Bayesian Brains without Probabilities

Adam N. Sanborn<sup>1,\*</sup> and Nick Chater<sup>2</sup>

Bayesian explanations have swept through cognitive science over the past two decades, from intuitive physics and causal learning, to perception, motor con-

<sup>2</sup>Division of Psychology and Language Sciences, University College London, Gower Street, London WC1E 6BT, UK  
<sup>3</sup>Department of Psychology, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh PA 15213, USA  
<sup>4</sup>School of Psychology, University of Adelaide, Mawson Lakes, Adelaide, SA 5005, Australia  
<sup>5</sup>Brain and Cognitive Sciences Department, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

**Cognitive science aims to reverse-engineer the mind, and many of the engineering challenges the mind faces** with abstract principles that allow agents to solve problems posed by the world – the functions that n

Psychological Review

© 2017 American Psychological Association  
0033-295X/17/\$12.00 <http://dx.doi.org/10.1037/rev0000052>

## Bayesian Models of Cognition Revisited: Setting Optimality Aside and Letting Data Drive Psychological Theory

Sean Taylor and Daniel J. Navarro

Amy Perfors

Psychological Review  
2014, Vol. 121, No. 3, 463–480

© 2014 American Psychological Association  
0033-295X/14/\$12.00 <http://dx.doi.org/10.1037/a0037010>

## Surprisingly Rational: Probability Theory Plus Noise Explains Biases in Judgment

Fintan Costello  
University College Dublin

Paul Watts  
National University of Ireland

The systematic biases seen in people's probability judgments are typically taken as evidence that people do not use the rules of probability theory when reasoning about probability but instead use heuristics, which sometimes yield reasonable judgments and sometimes yield systematic biases. This view has had a major impact in economics, law, medicine, and other fields; indeed, the idea that people cannot reason with probabilities has become a truism. We present a simple alternative to this view, where people reason about probability according to probability theory but are subject to random variation or noise in the reasoning process. In this account the effect of noise is canceled for some probabilistic expressions. Analyzing data from 2 experiments, we find that, for these expressions, people's probability judgments are strikingly close to those required by probability theory. For other expressions, this account produces systematic deviations in probability estimates. These deviations explain 4 reliable biases in human probabilistic reasoning (conservatism, subadditivity, conjunction, and disjunction fallacies). These results suggest that people's probability judgments embody the rules of probability theory and that biases in those judgments are due to the effects of random noise.

**Keywords:** probability, rationality, random variation, heuristics, biases

Surprisingly rational: Probability theory  
plus noise explains biases in judgment.

Costello, Fintan, Watts, Paul

Psychological Review, Vol 121(3), Jul 2014, 463-480



# The Mind is a **Bayesian Sampler**

Trends in Cognitive Sciences  
2016

**Opinion**

Bayesian Brains without  
Probabilities

Adam N. Sanborn<sup>1,\*</sup> and Nick Chater<sup>2</sup>

Surprisingly rational: Probability theory plus noise explains biases in judgment.

Costello, Fintan, Watts, Paul

Psychological Review, Vol 121(3), Jul 2014, 463-480

Brain does not represent probability distributions



## The Mind is a Bayesian Sampler

Trends in Cognitive Sciences  
2016

**Opinion**

Bayesian Brains without Probabilities

Adam N. Sanborn<sup>1,\*</sup> and Nick Chater<sup>2</sup>



Surprisingly rational: Probability theory  
plus noise explains biases in judgment.

Costello, Fintan, Watts, Paul

Psychological Review, Vol 121(3), Jul 2014, 463-480

Brain does not represent  
probability distributions

Probability distributions  
approximated via sampling  
from memory



# The Mind is a **Bayesian Sampler**

Trends in Cognitive Sciences  
2016

**Opinion**

Bayesian Brains without  
Probabilities

Adam N. Sanborn<sup>1,\*</sup> and Nick Chater<sup>2</sup>

Surprisingly rational: Probability theory  
plus noise explains biases in judgment.

Costello, Fintan, Watts, Paul

Psychological Review, Vol 121(3), Jul 2014, 463-480

Brain does not represent  
probability distributions

Probability distributions  
approximated via sampling  
from memory



## The Mind is a **Bayesian Sampler**

*Test Bayesian brain:*  
Can sampling based  
Bayesian model describe  
participants' behavior?

Trends in Cognitive Sciences  
2016

### Opinion

## Bayesian Brains without Probabilities

Adam N. Sanborn<sup>1,\*</sup> and Nick Chater<sup>2</sup>

Surprisingly rational: Probability theory plus noise explains biases in judgment.

Costello, Fintan, Watts, Paul

Psychological Review, Vol 121(3), Jul 2014, 463-480



# The Mind is a Bayesian Sampler

Trends in Cognitive Sciences  
2016

## Opinion

Bayesian Brains without Probabilities

Adam N. Sanborn<sup>1,\*</sup> and Nick Chater<sup>2</sup>

Brain does not represent probability distributions

Probability distributions approximated via sampling from memory

*Test Bayesian brain:*  
Can sampling based Bayesian model describe participants' behavior?

*Noisy sampler:*  
Model for probability estimation, combines

- Bayesian brain
- Sampling
- Noise

Surprisingly rational: Probability theory plus noise explains biases in judgment.

Costello, Fintan, Watts, Paul

Psychological Review, Vol 121(3), Jul 2014, 463-480



## The Mind is a Bayesian Sampler

Trends in Cognitive Sciences  
2016

### Opinion

Bayesian Brains without Probabilities

Adam N. Sanborn<sup>1,\*</sup> and Nick Chater<sup>2</sup>

Brain does not represent probability distributions

Probability distributions approximated via sampling from memory

*Test Bayesian brain:*  
Can sampling based Bayesian model describe participants' behavior?

*Noisy sampler:*  
Model for probability estimation, combines

- Bayesian brain
- Sampling
- Noise

Kind of.

Surprisingly rational: Probability theory  
plus noise explains biases in judgment.

Costello, Fintan, Watts, Paul

Psychological Review, Vol 121(3), Jul 2014, 463-480



# The Mind is a **Bayesian Sampler**

Trends in Cognitive Sciences  
2016

**Opinion**

Bayesian Brains without  
Probabilities

Adam N. Sanborn<sup>1,\*</sup> and Nick Chater<sup>2</sup>

Surprisingly rational: Probability theory plus noise explains biases in judgment.

Costello, Fintan, Watts, Paul

Psychological Review, Vol 121(3), Jul 2014, 463-480

What is the probability that the weather will be *cloudy* on a randomly selected day in Ireland?



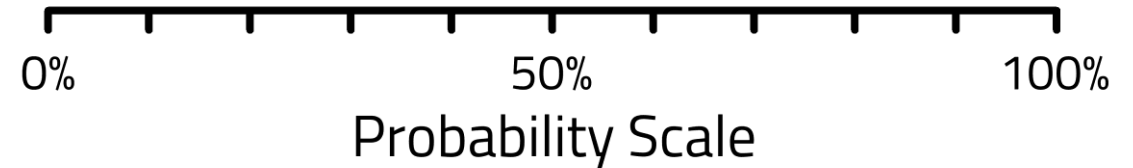
## The Mind is a Bayesian Sampler

Trends in Cognitive Sciences  
2016

**Opinion**

Bayesian Brains without Probabilities

Adam N. Sanborn<sup>1,\*</sup> and Nick Chater<sup>2</sup>





Surprisingly rational: Probability theory plus noise explains biases in judgment.

Costello, Fintan, Watts, Paul

Psychological Review, Vol 121(3), Jul 2014, 463-480



## The Mind is a Bayesian Sampler

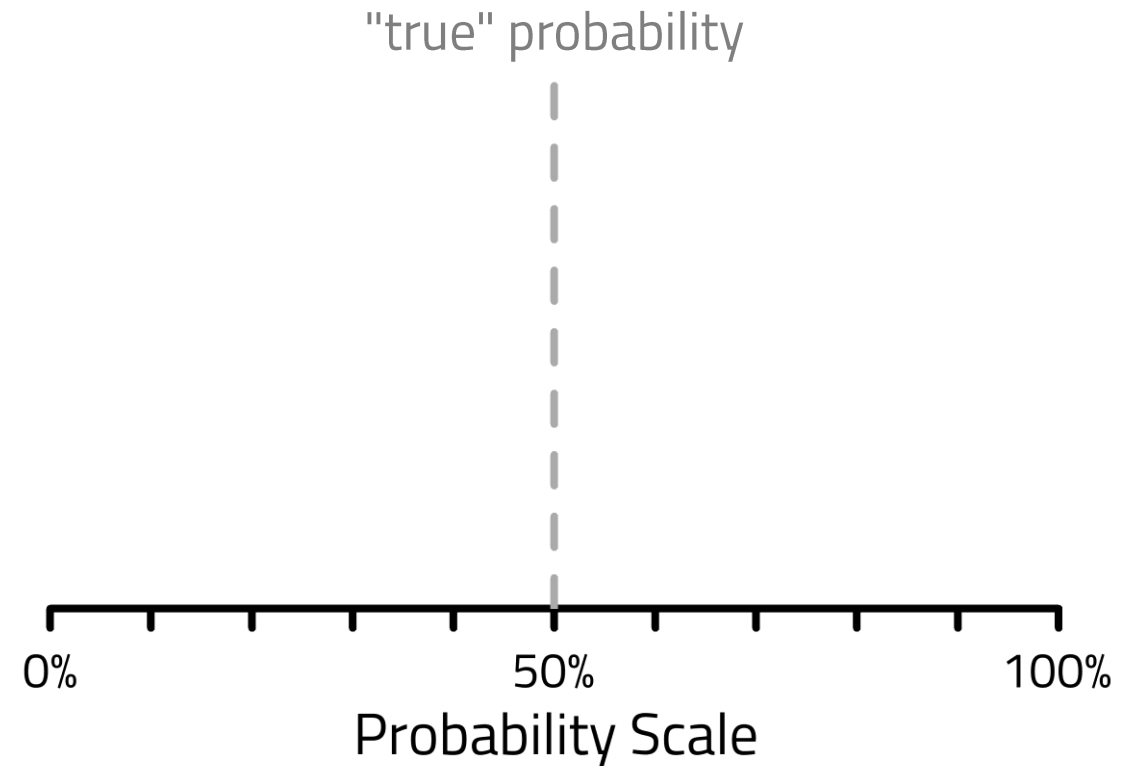
Trends in Cognitive Sciences  
2016

**Opinion**

Bayesian Brains without Probabilities

Adam N. Sanborn<sup>1,\*</sup> and Nick Chater<sup>2</sup>

What is the probability that the weather will be *cloudy* on a randomly selected day in Ireland?



Surprisingly rational: Probability theory plus noise explains biases in judgment.

Costello, Fintan, Watts, Paul

Psychological Review, Vol 121(3), Jul 2014, 463-480



## The Mind is a Bayesian Sampler

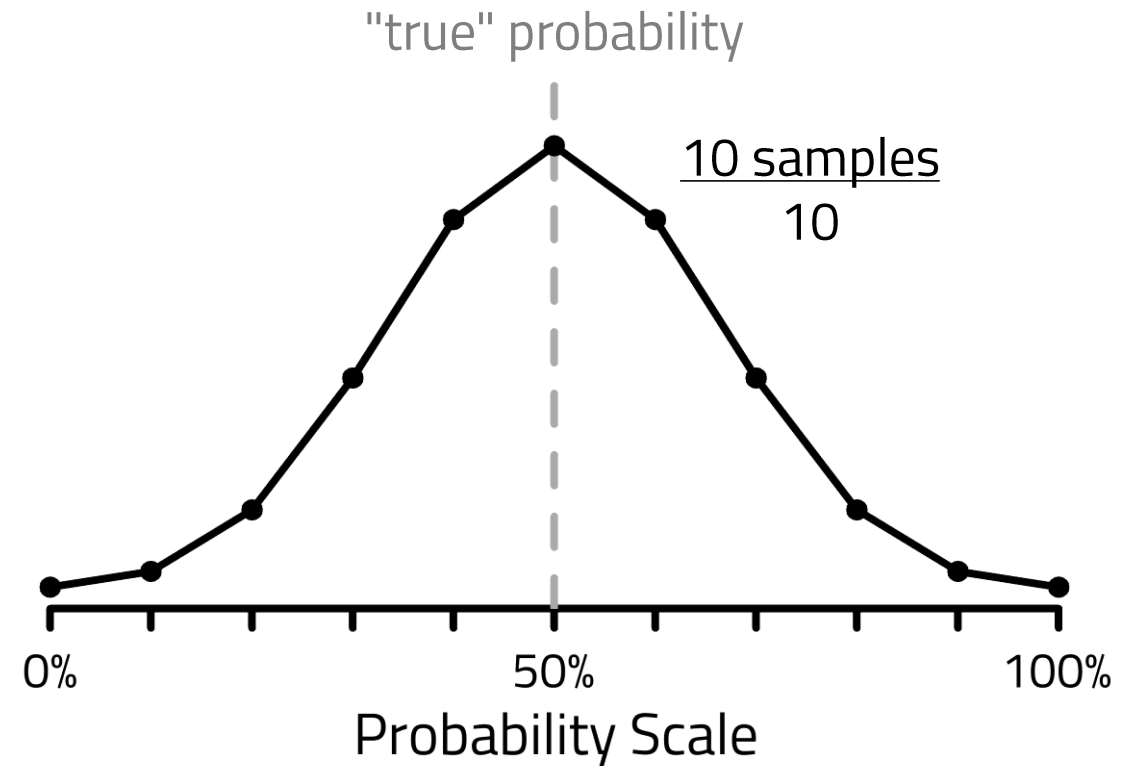
Trends in Cognitive Sciences  
2016

**Opinion**

Bayesian Brains without Probabilities

Adam N. Sanborn<sup>1,\*</sup> and Nick Chater<sup>2</sup>

What is the probability that the weather will be *cloudy* on a randomly selected day in Ireland?



Surprisingly rational: Probability theory plus noise explains biases in judgment.

Costello, Fintan, Watts, Paul

Psychological Review, Vol 121(3), Jul 2014, 463-480



## The Mind is a Bayesian Sampler

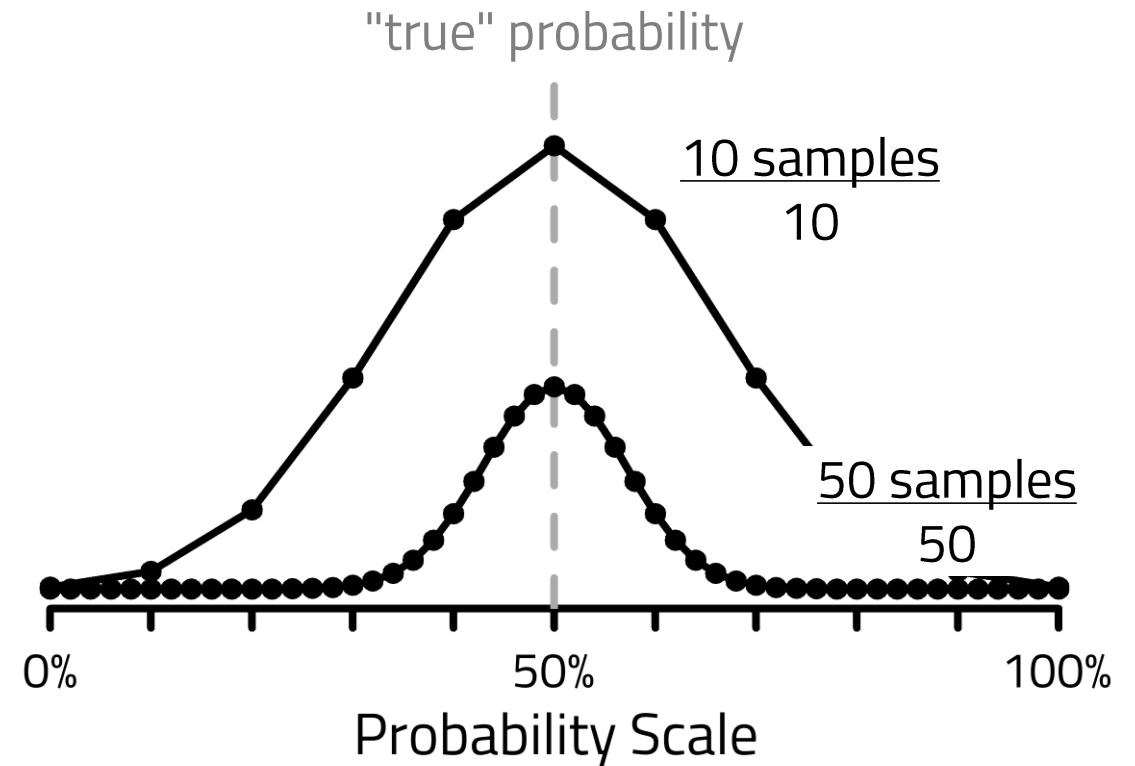
Trends in Cognitive Sciences  
2016

**Opinion**

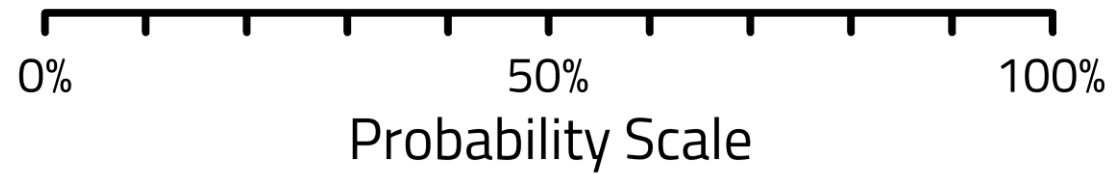
Bayesian Brains without Probabilities

Adam N. Sanborn<sup>1,\*</sup> and Nick Chater<sup>2</sup>

What is the probability that the weather will be *cloudy* on a randomly selected day in Ireland?

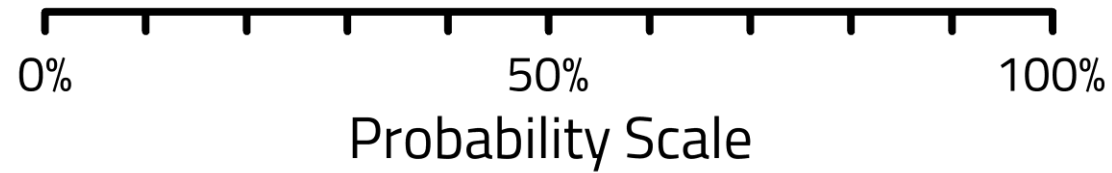


What is the probability that the weather will be ... on a randomly selected day in Ireland?



01 | *cloudy?*

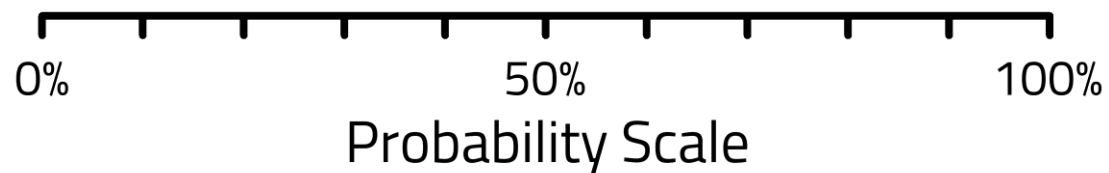
What is the probability that the weather will be ... on a randomly selected day in Ireland?



01 | *cloudy?*

02 | *cold?*

What is the probability that the weather will be ... on a randomly selected day in Ireland?



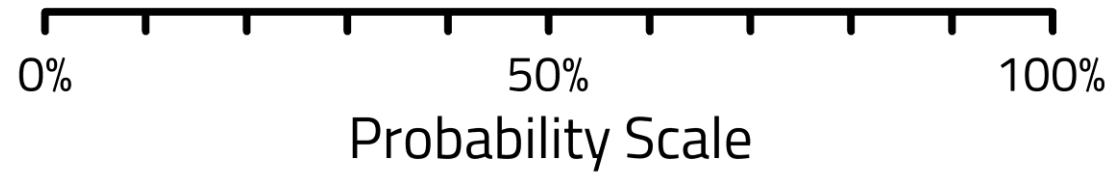
01 | *cloudy?*

02 | *cold?*

03 | *cloudy or  
cold?*



What is the probability that the weather will be ... on a randomly selected day in Ireland?



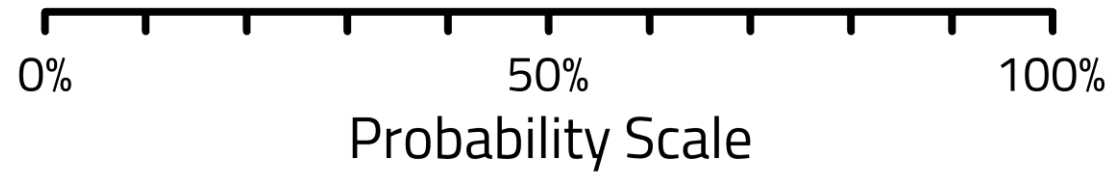
01 | *cloudy?*

02 | *cold?*

03 | *cloudy or  
cold?*

04 | *cloudy and  
cold?*

What is the probability that the weather will be ... on a randomly selected day in Ireland?



01 | *cloudy?*

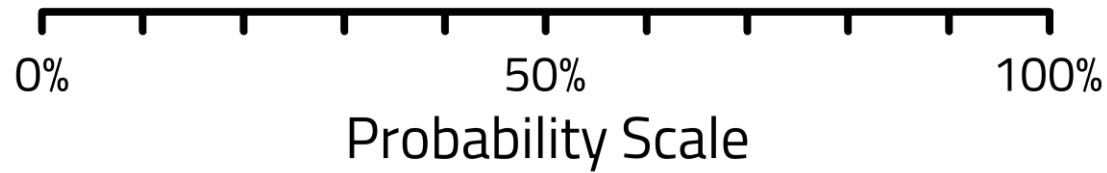
02 | *cold?*

03 | *cloudy or cold?*

04 | *cloudy and cold?*

05 | *cloudy and not cold?*

What is the probability that the weather will be ... on a randomly selected day in Ireland?



01 | *cloudy?*

02 | *cold?*

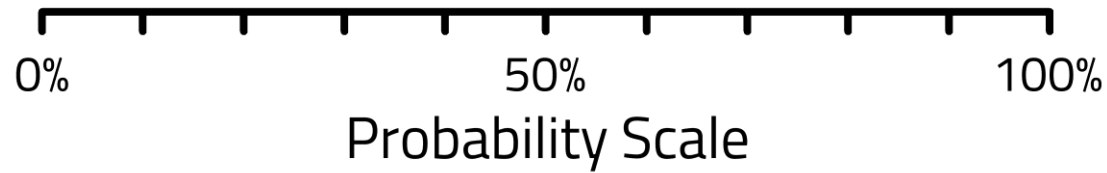
03 | *cloudy or cold?*

04 | *cloudy and cold?*

05 | *cloudy and not cold?*

06 | *cold and not cloudy?*

What is the probability that the weather will be ... on a randomly selected day in Ireland?



01

*cloudy?*

02

*cold?*

03

*cloudy or  
cold?*

04

*cloudy and  
cold?*

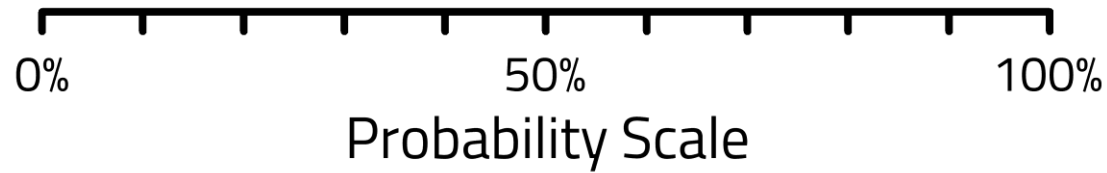
05

*cloudy and  
not cold?*

06

*cold and  
not cloudy?*

What is the probability that the weather will be ... on a randomly selected day in Ireland?



01 | *cloudy?*

02 | *cold?*

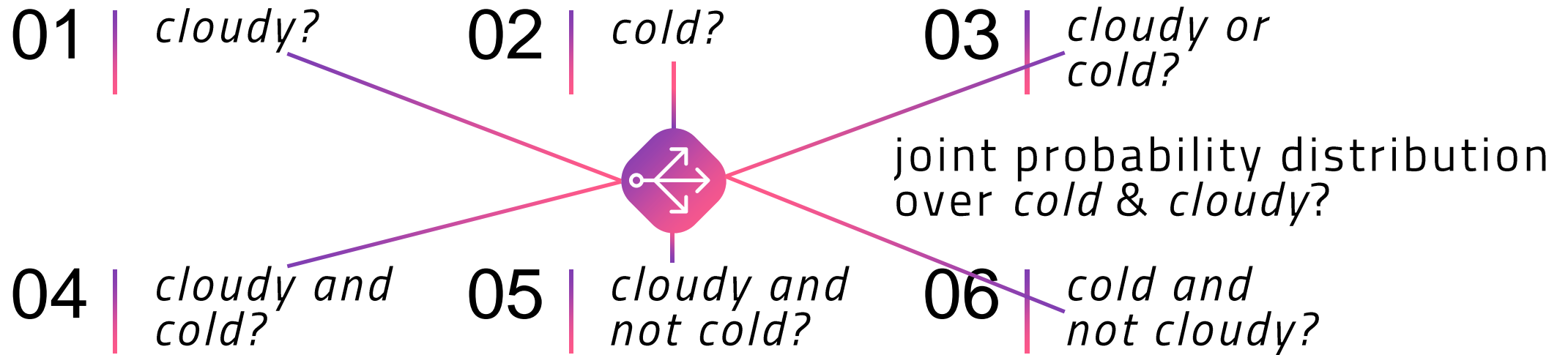
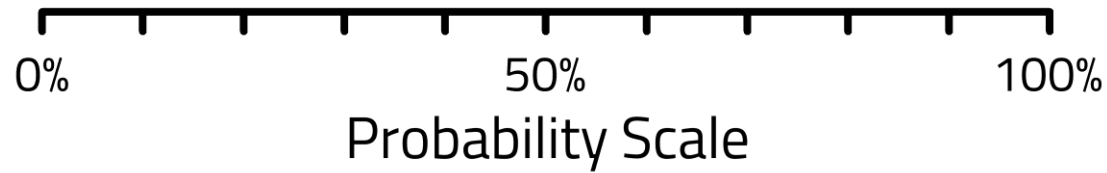
03 | *cloudy or cold?*

04 | *cloudy and cold?*

05 | *cloudy and not cold?*

06 | *cold and not cloudy?*

What is the probability that the weather will be ... on a randomly selected day in Ireland?



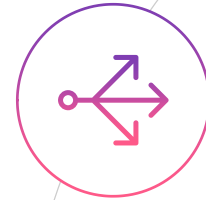


# Noisy Sampler Generative Model for Probability Estimation

---

## Joint Probability Distribution

Provides "true" probability  
(e.g., Oakford & Chater, 2014)



## Bayesian Sampling

Probabilistic transformation of  
"true" probability



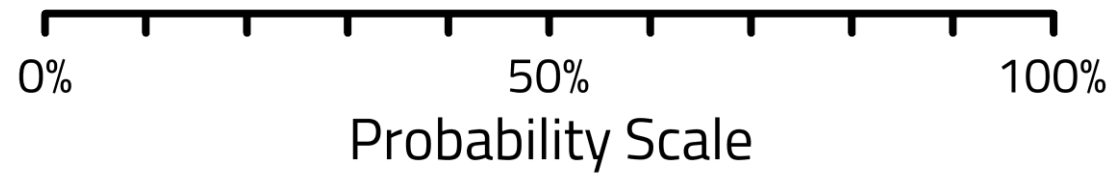
## Memory & Response Noise

Sometimes sampling is erroneous  
(Costello & Watts, 2014) *and*  
sometimes response is noisy



Noisy Sampler

# Discrete Probability Distribution on

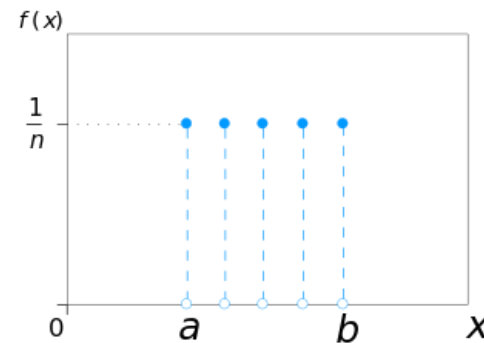


Mixture Model

$$\hat{P}(A) = \frac{\text{Noisy number of } A \text{ instances}}{\text{Total number of samples}}$$

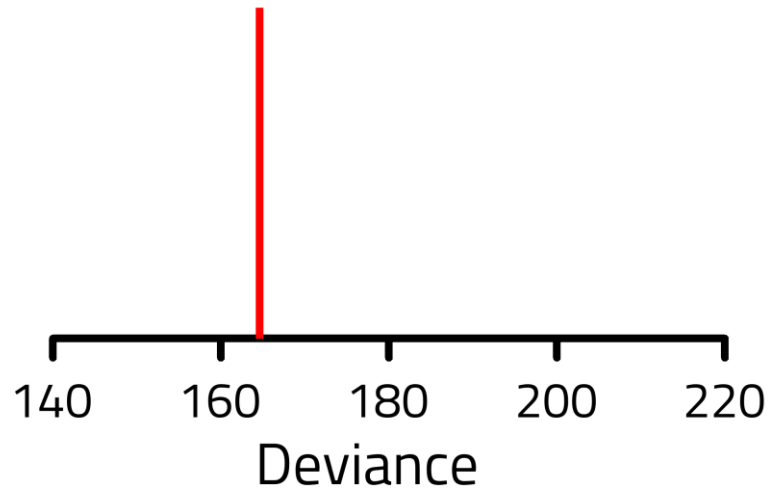
$(1 - \phi)$

$\phi$



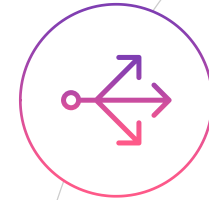
Discrete uniform  
distribution to  
adjacent points

# Noisy Sampler Generative Model for Probability Estimation



## Joint Probability Distribution

Provides "true" probability  
(e.g., Oakford & Chater, 2014)



## Bayesian Sampling

Probabilistic transformation of  
"true" probability



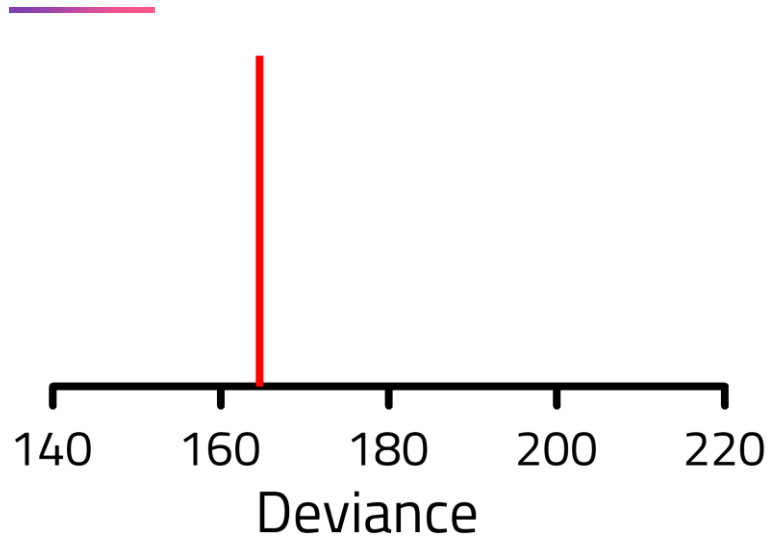
## Memory & Response Noise

Sometimes sampling is erroneous  
(Costello & Watts, 2014) *and*  
sometimes response is noisy

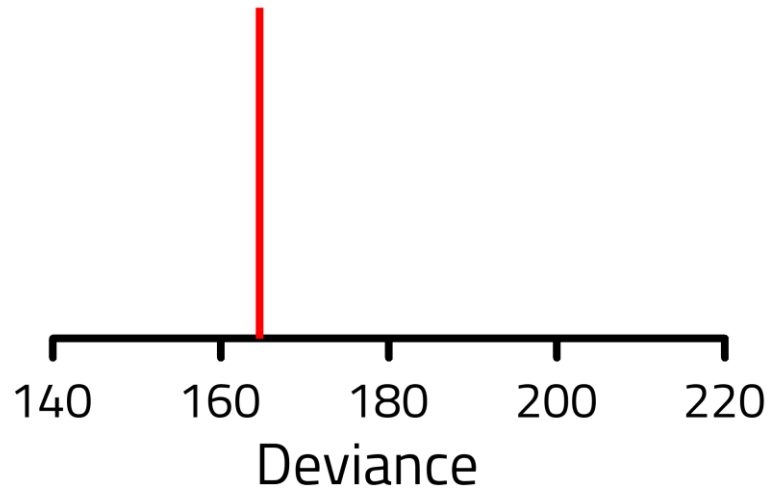


Derive distribution of deviance  
under null hypothesis ( $H_0$ )

*Noisy Sampler*  
Generative Model  
for Probability  
Estimation



*Noisy Sampler*  
Generative Model  
for Probability  
Estimation

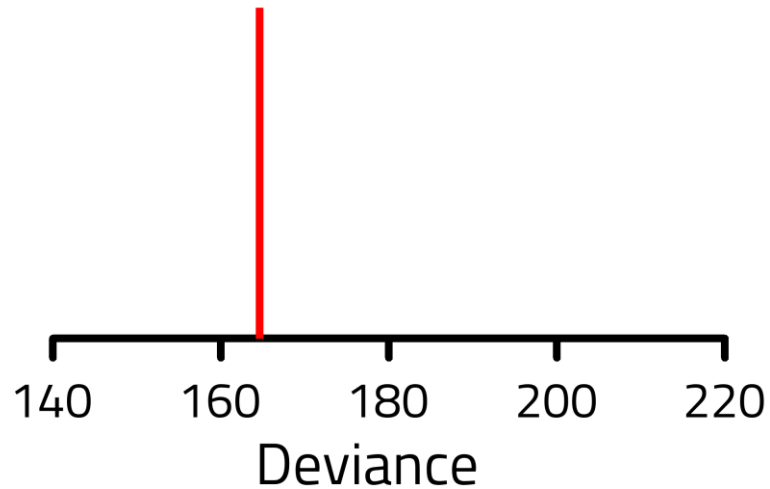


Derive distribution of deviance  
under null hypothesis ( $H_0$ )



Generate synthetic  
data from best fitting  
parameter estimates

# Noisy Sampler Generative Model for Probability Estimation



Derive distribution of deviance  
under null hypothesis ( $H_0$ )



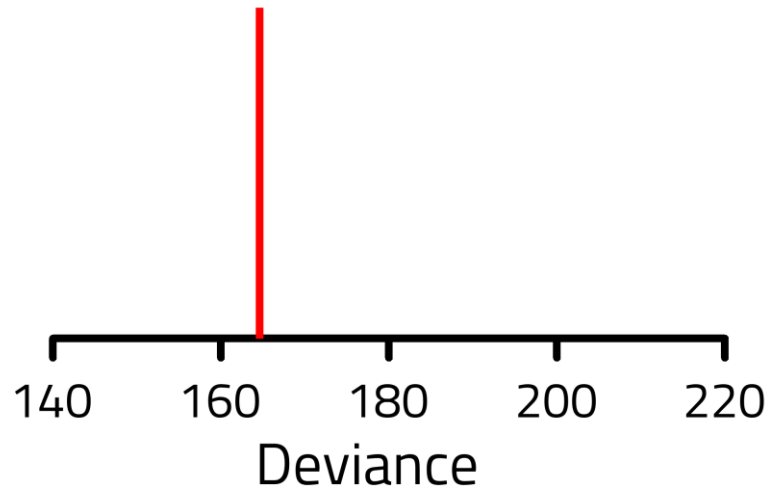
Generate synthetic  
data from best fitting  
parameter estimates



Fit *noisy sampler* to  
synthetic data



## Noisy Sampler Generative Model for Probability Estimation



Derive distribution of deviance  
under null hypothesis ( $H_0$ )



Generate synthetic  
data from best fitting  
parameter estimates

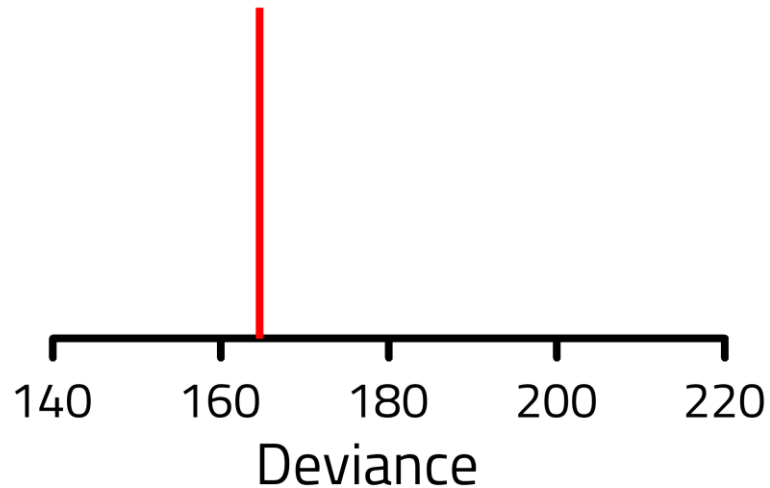


Fit *noisy sampler* to  
synthetic data



Record deviance of fit  
to synthetic data

## Noisy Sampler Generative Model for Probability Estimation



Derive distribution of deviance  
under null hypothesis ( $H_0$ )



Generate synthetic  
data from best fitting  
parameter estimates



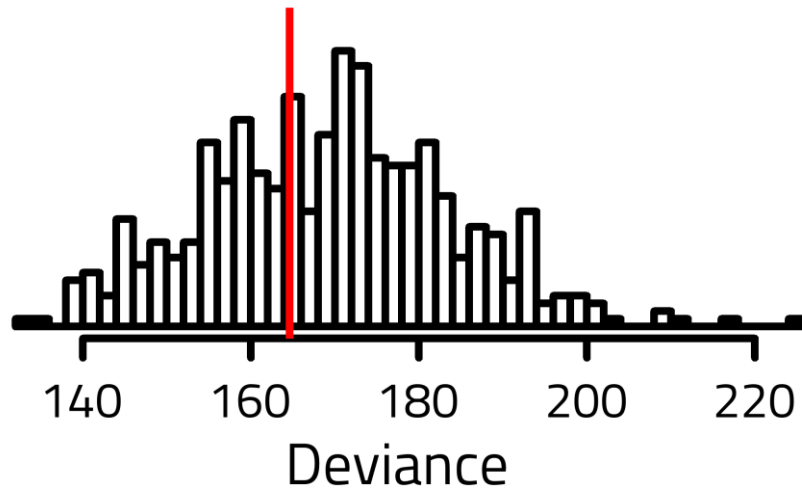
Fit *noisy sampler* to  
synthetic data



Record deviance of fit  
to synthetic data

repeat  
500  
times

## Noisy Sampler Generative Model for Probability Estimation



Derive distribution of deviance  
under null hypothesis ( $H_0$ )

01

Generate synthetic  
data from best fitting  
parameter estimates

02

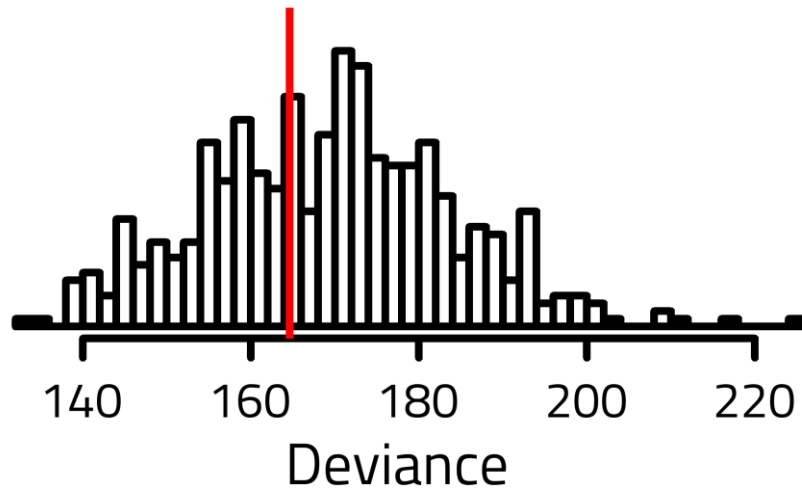
Fit *noisy sampler* to  
synthetic data

03

Record deviance of fit  
to synthetic data

repeat  
500  
times

## Noisy Sampler Generative Model for Probability Estimation



Derive distribution of deviance  
under null hypothesis ( $H_0$ )

01

Generate synthetic  
data from best fitting  
parameter estimates

02

Fit *noisy sampler* to  
synthetic data

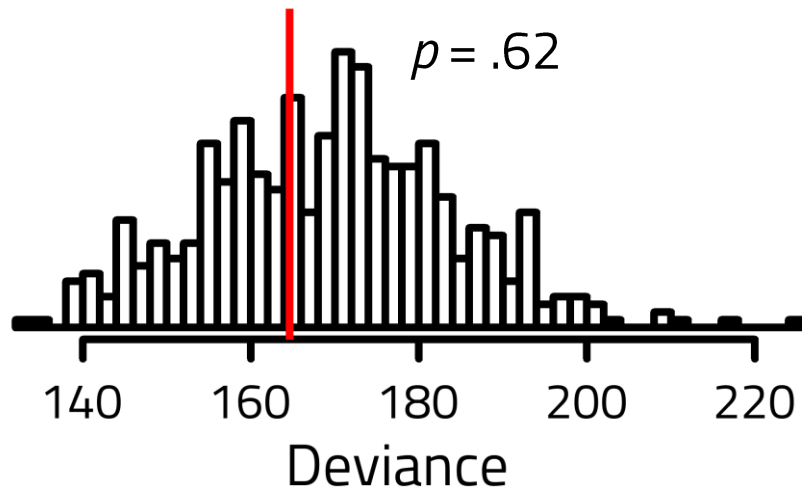
03

Record deviance of fit  
to synthetic data

repeat  
500  
times

*p*-value: probability of sampling  
test statistic at least as  
extreme as observed under  $H_0$

## Noisy Sampler Generative Model for Probability Estimation



Derive distribution of deviance  
under null hypothesis ( $H_0$ )



Generate synthetic  
data from best fitting  
parameter estimates



Fit *noisy sampler* to  
synthetic data



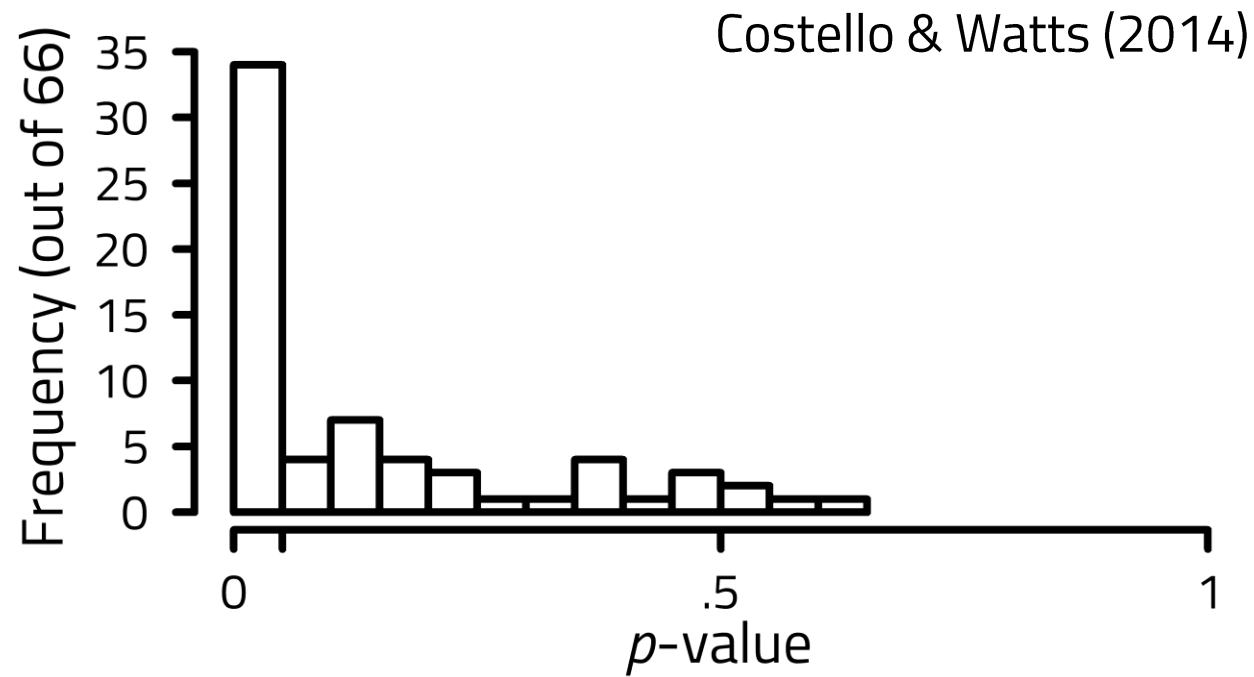
Record deviance of fit  
to synthetic data

repeat  
500  
times

$p$ -value: probability of sampling  
test statistic at least as  
extreme as observed under  $H_0$

1.

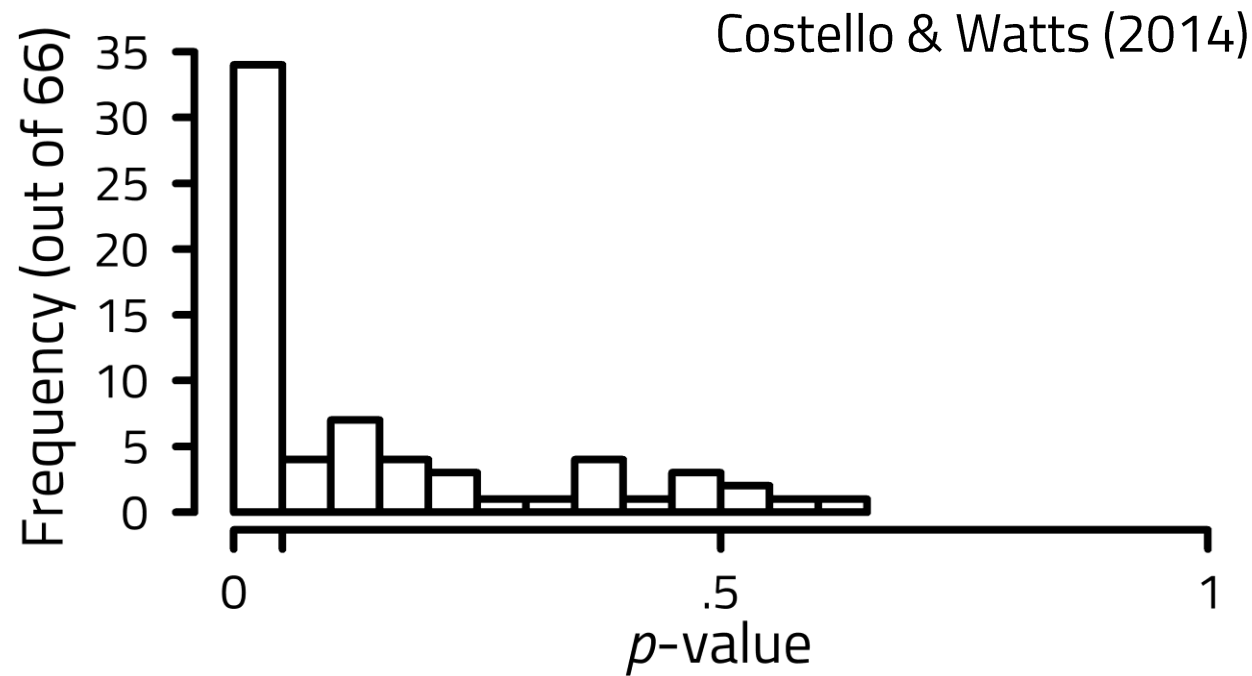
Testing the Bayesian brain



1.

Testing the Bayesian brain

50% of participants  
Bayesian according to  
*noisy sampler*.



1.

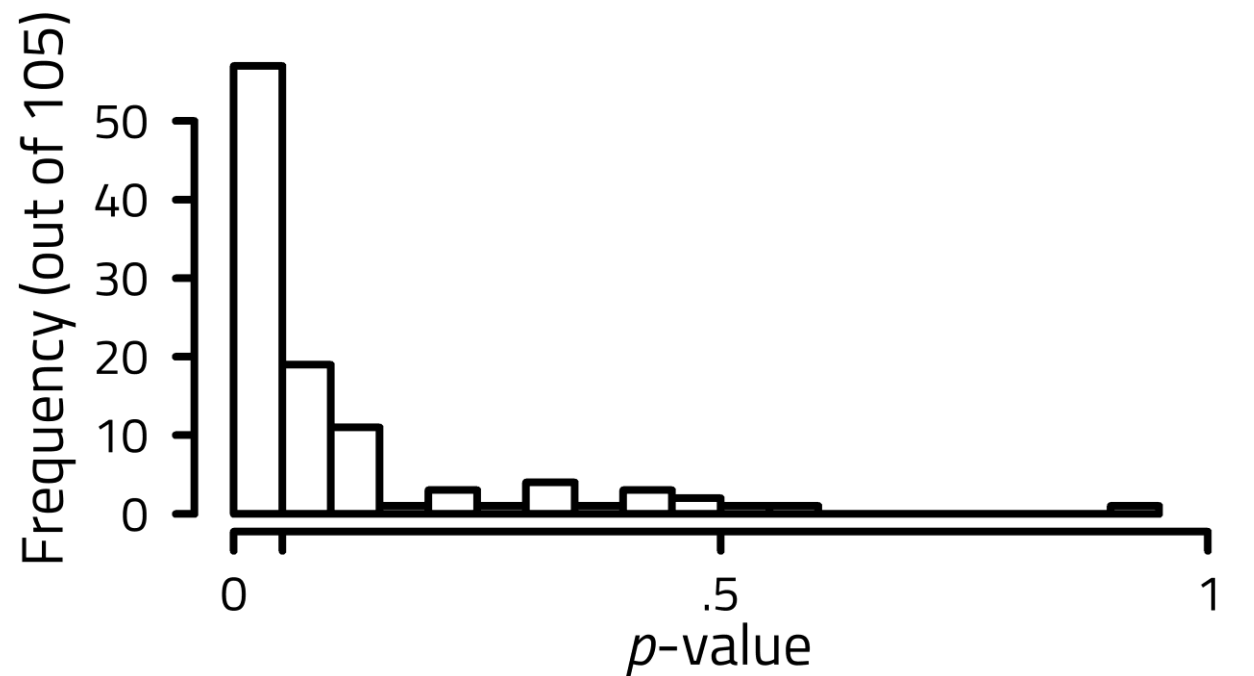
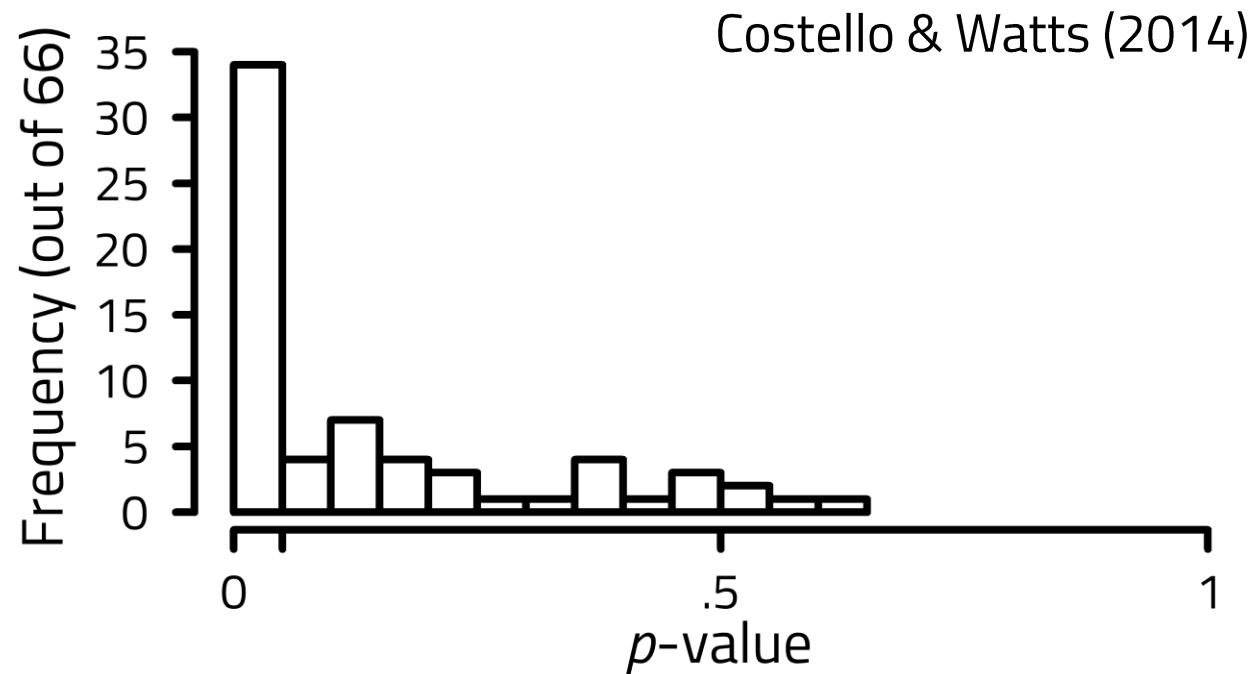
Testing the Bayesian brain

50% of participants  
Bayesian according to  
*noisy sampler*.

2.

MTurk online replication

46% of participants  
Bayesian according to  
*noisy sampler*.

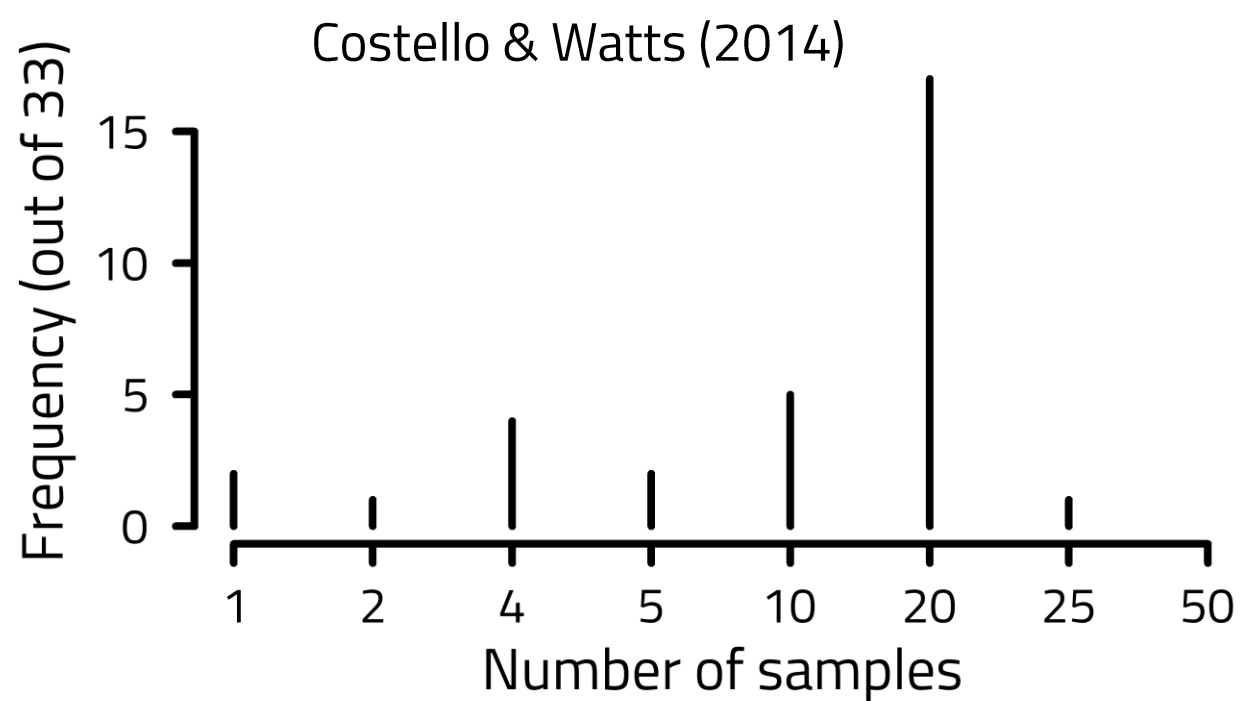




1.

How many samples do individuals take?

Restricted to *non-rejected* participants.



1.

How many samples do individuals take?



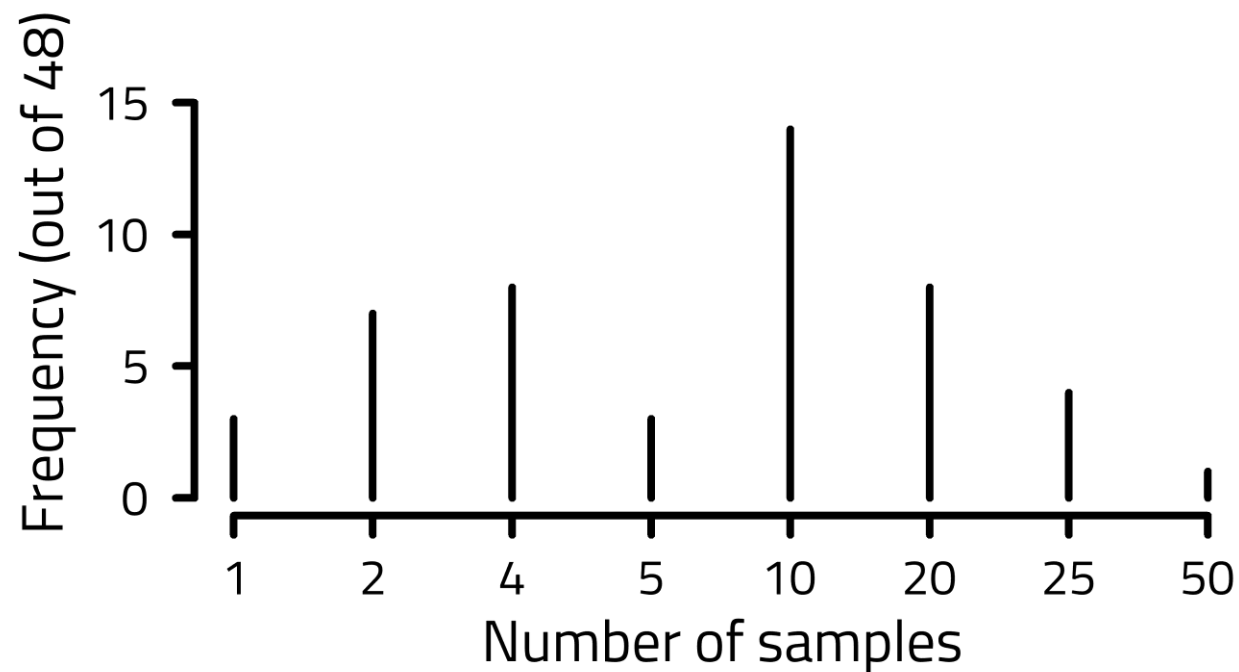
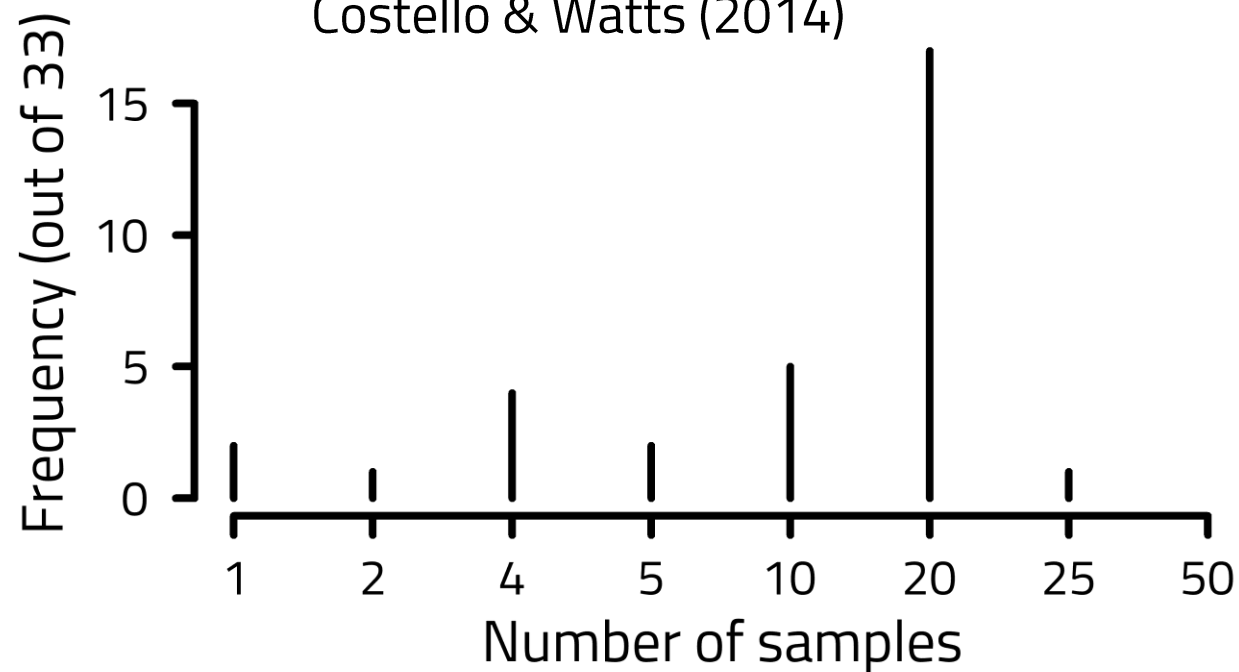
Restricted to *non-rejected* participants.


2.

MTurk online replication



Costello & Watts (2014)





Does sampling save  
the Bayesian brain?



THANK YOU FOR  
YOUR ATTENTION!

Questions?

cold – cloudy

cold – sunny

windy – cold

cloudy – rainy

cloudy – icy

sunny – icy

sunny – rainy

icy – windy

rainy – windy

= 42 independent data points per participant

Costello & Watts (2014)

## MODELING

---

- 19 for joint probability distributions
- 1 mixture weight
- 1 noise parameter
- 1 number of samples

cold – cloudy

cold – sunny

windy – cold

cloudy – rainy

cloudy – icy

sunny – icy

sunny – rainy

icy – windy

rainy – windy

= 42 independent data points per participant

Costello & Watts (2014)