Lab this Thursday?

Why a static prior is too much of a good thing

Bayesian modelers should incorporate more realistic memory theories

university of groningen

Hedderik van Rijn

Observation =

“Perception”

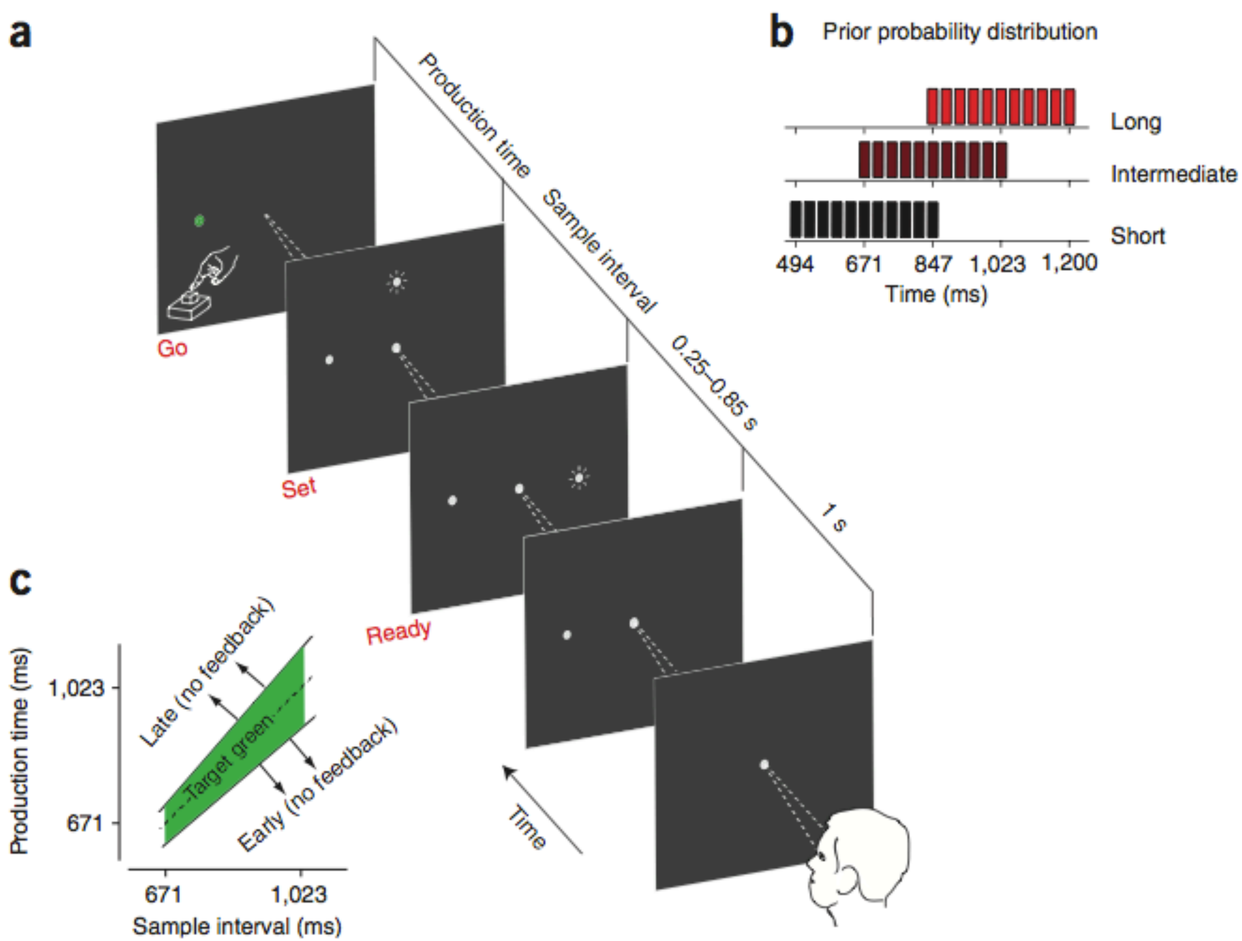
ARTICLES

nature neuroscience

Temporal context calibrates interval timing

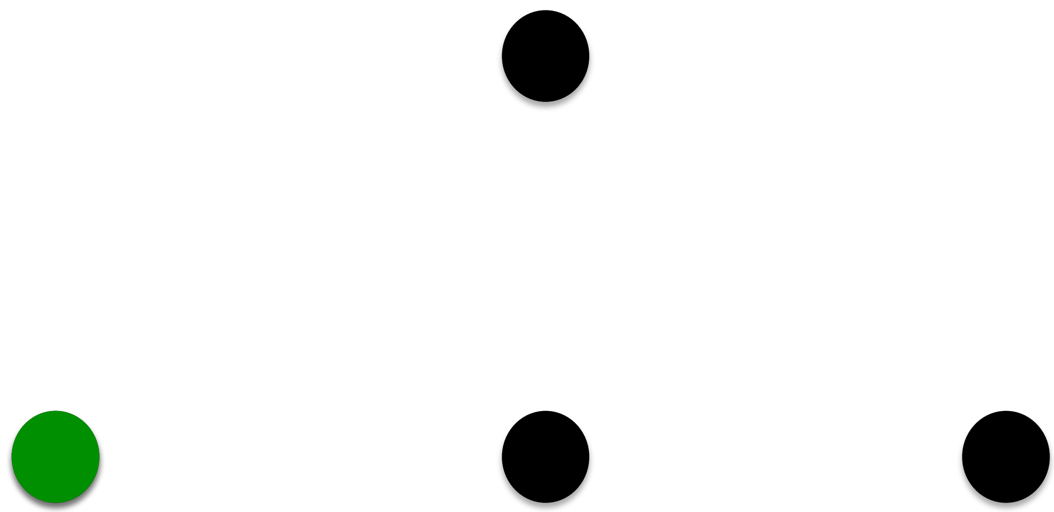
Mehrdad Jazayeri[1,2] & Michael N Shadlen[2]

Jazayeri & Shadlen, 2010

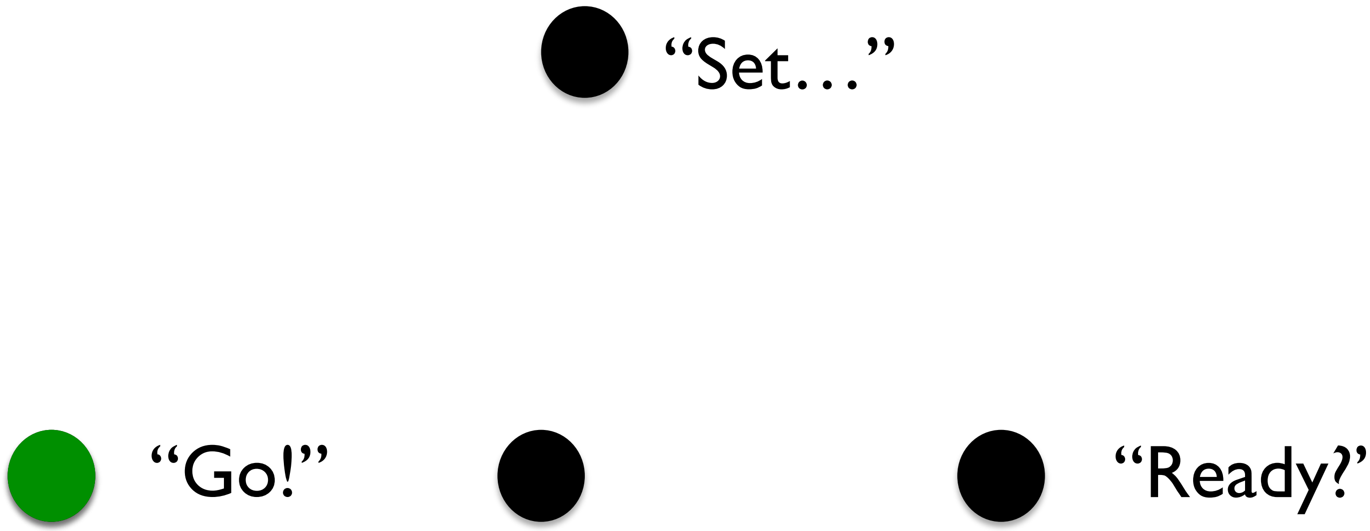


a
Production time
Sample interval
0.25–0.85 s
1 s
Go
Set
Ready
Time
b
Prior probability distribution
Long
Intermediate
Short
494
671
847
1,023
1,200
Time (ms)
c
Late (no feedback)
Target green
Early (no feedback)
Production time (ms)
1,023
671
671
1,023
Sample interval (ms)

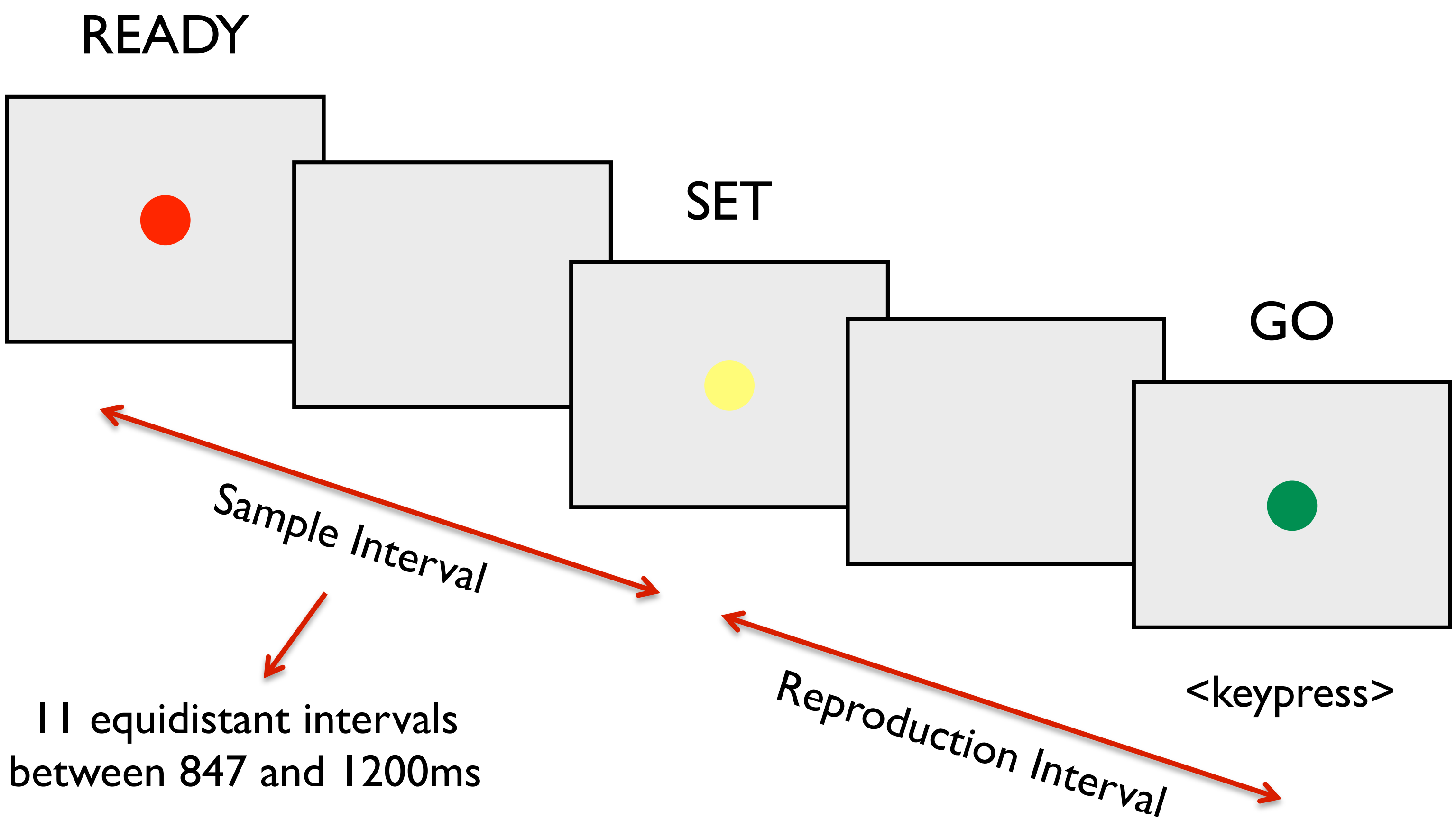
Interval Timing



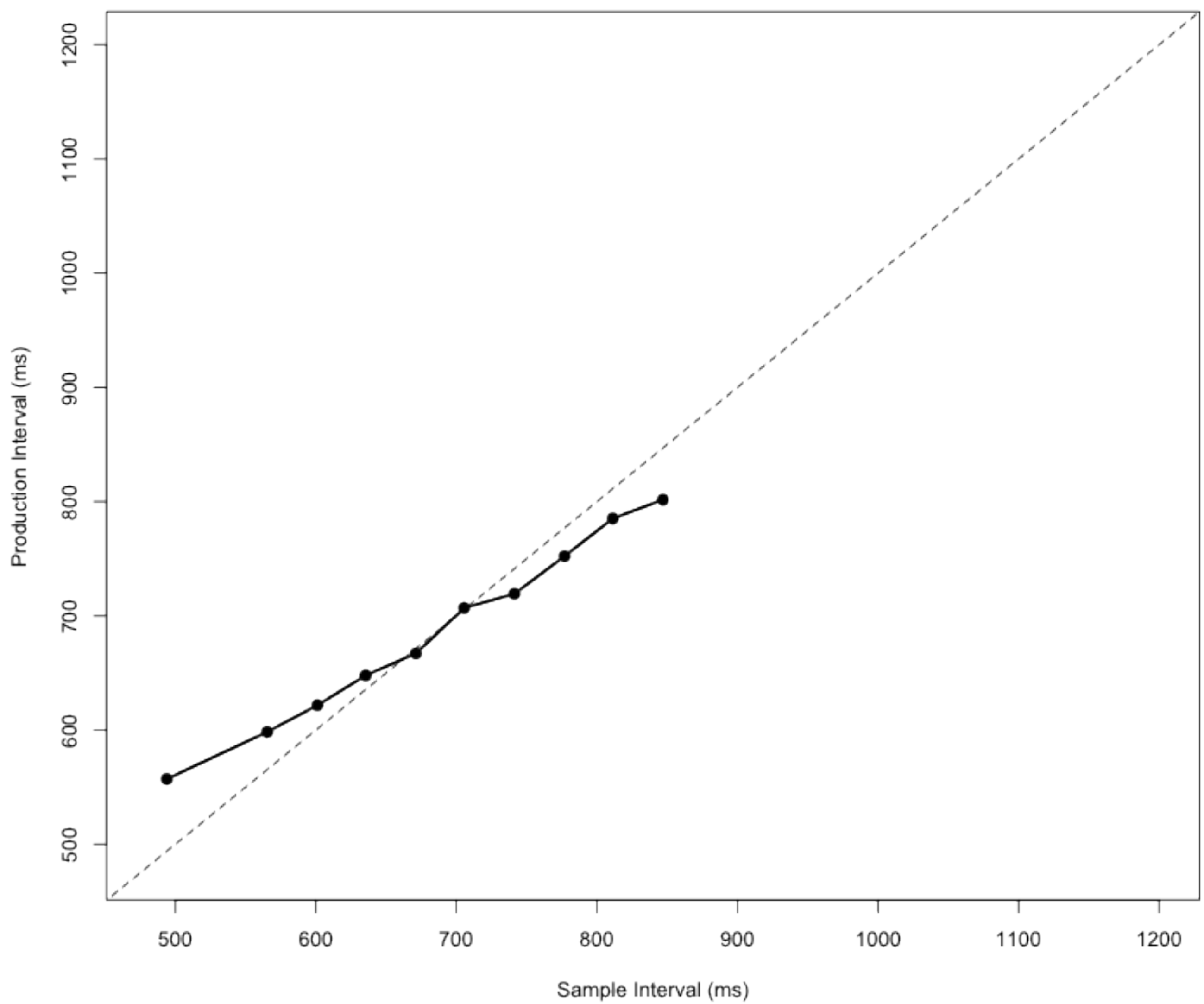
Interval Timing

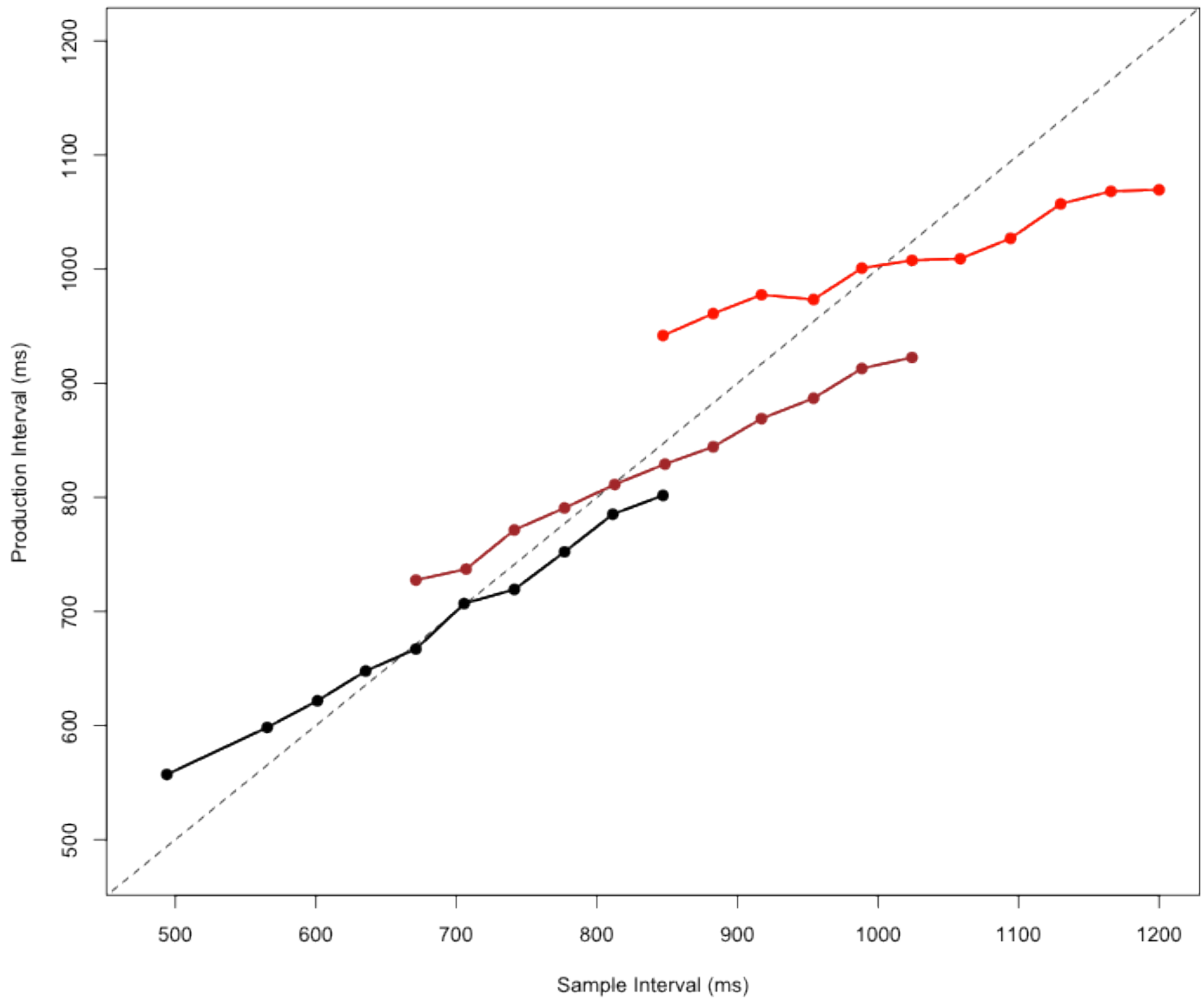


Ready-Set-Go paradigm

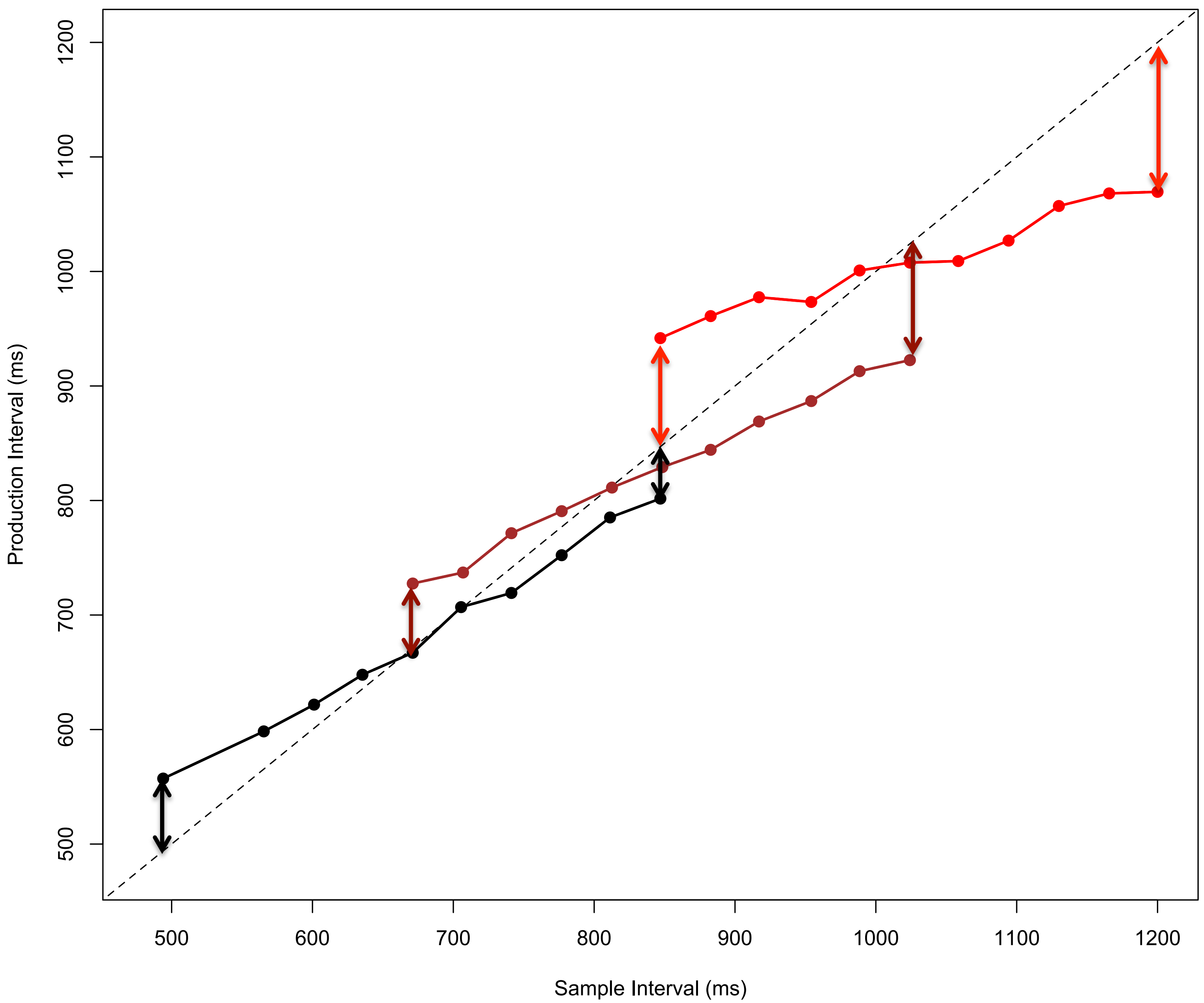


Jazayeri & Shadlen, 2010


Production Interval (ms)
500
600
700
800
900
1000
1100
1200
Sample Interval (ms)



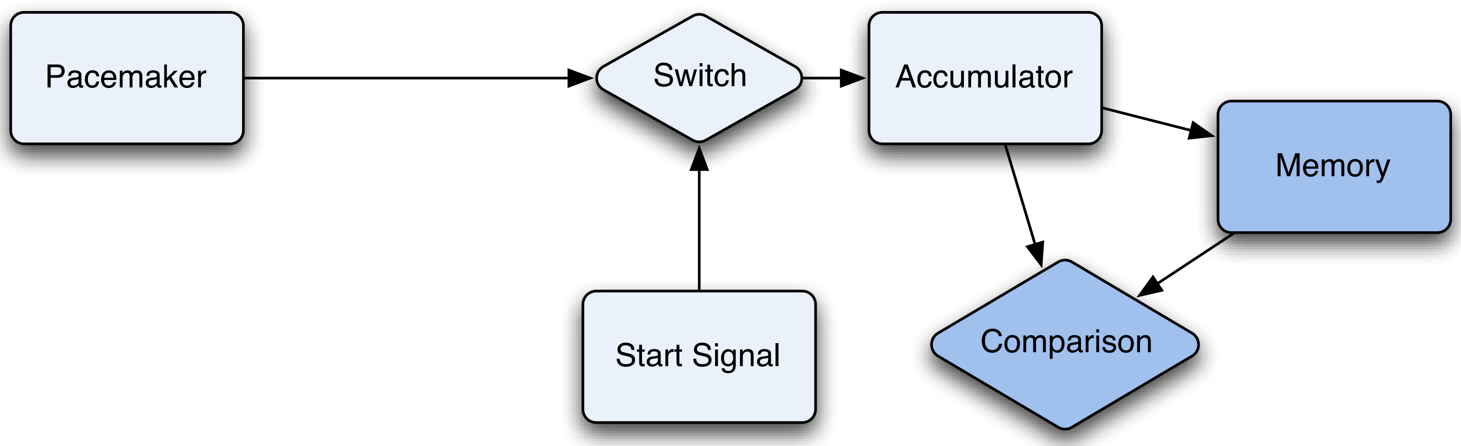
1200
1100
1000
900
800
700
600
500
Production Interval (ms)
500
600
700
800
900
1000
1100
1200
Sample Interval (ms)



Production Interval (ms)
500
600
700
800
900
1000
1100
1200
Sample Interval (ms)

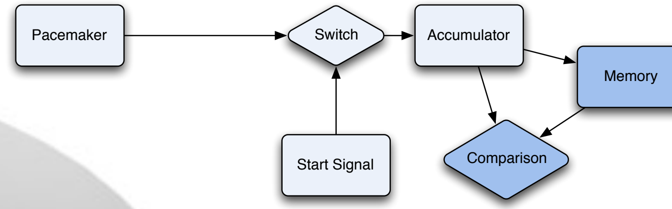
How to explain this data?

Information-Processing Models

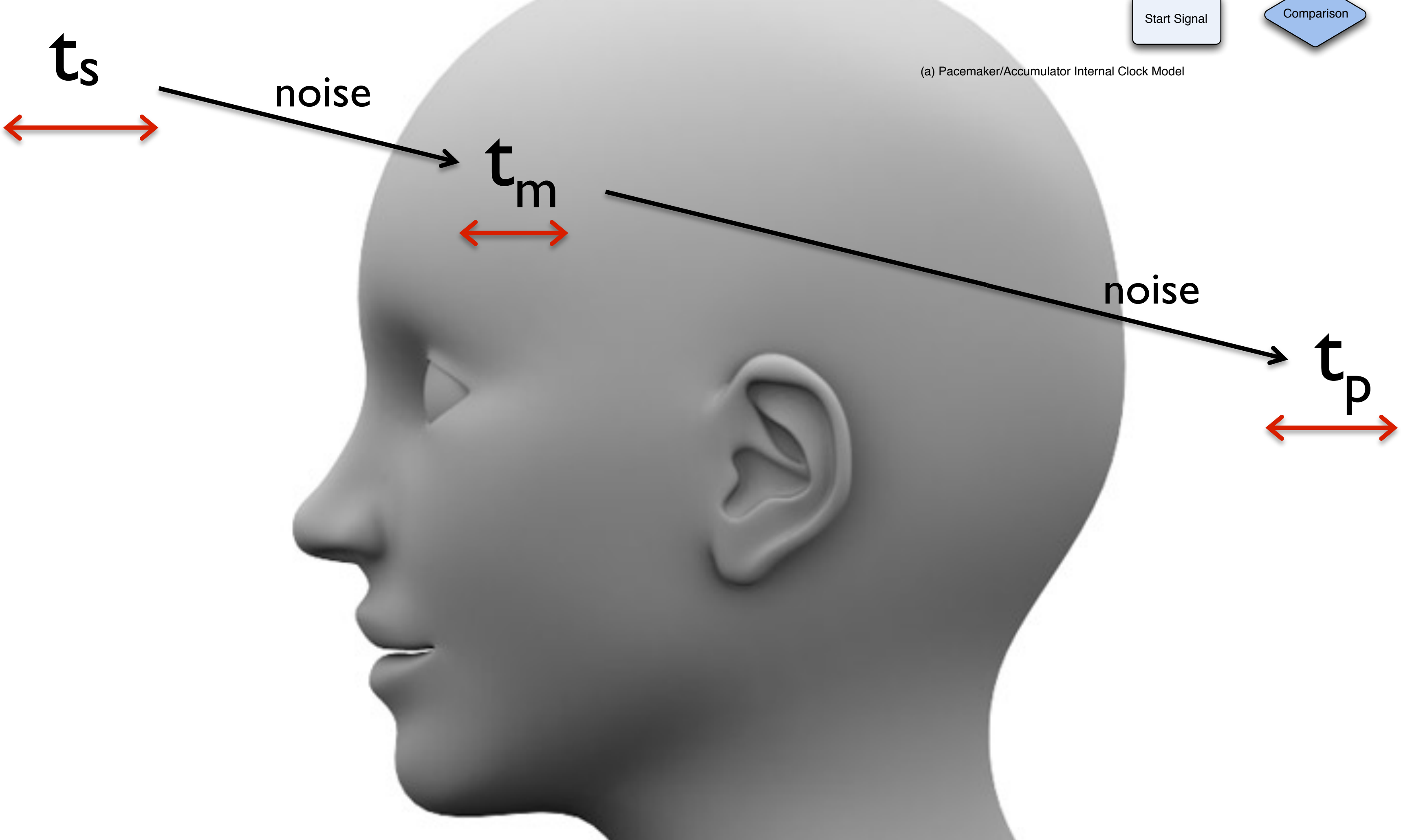


(a) Pacemaker/Accumulator Internal Clock Model

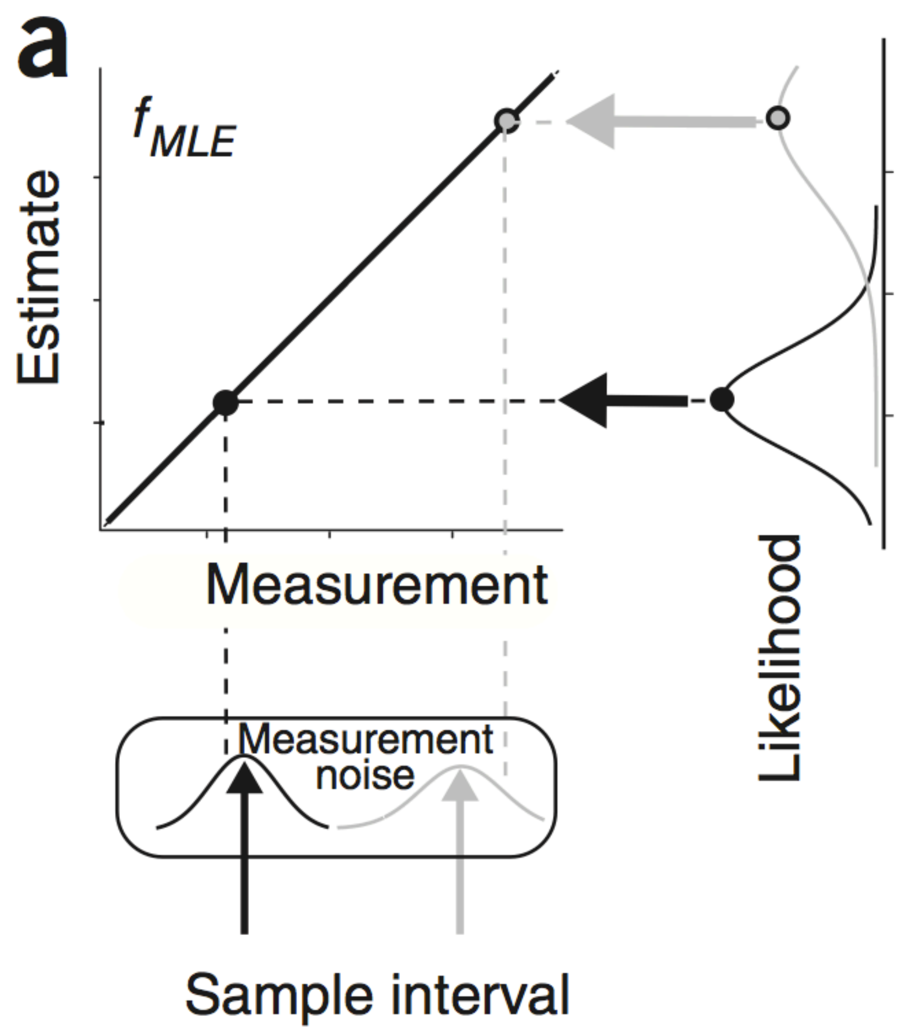
IP Models Reformulated



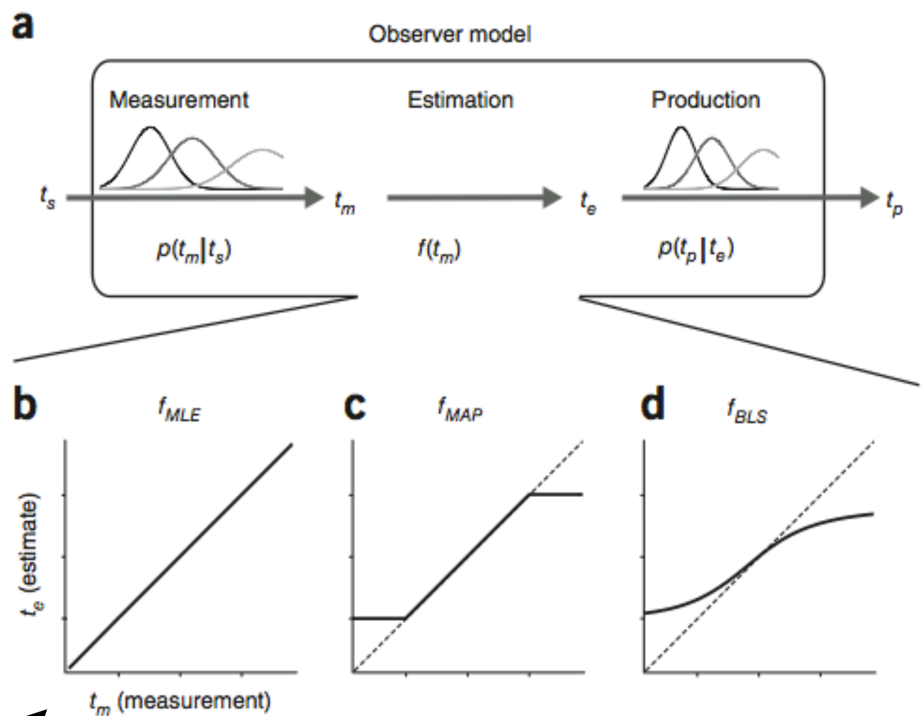
(a) Pacemaker/Accumulator Internal Clock Model



Maximum-likelihood estimation
“No Context”



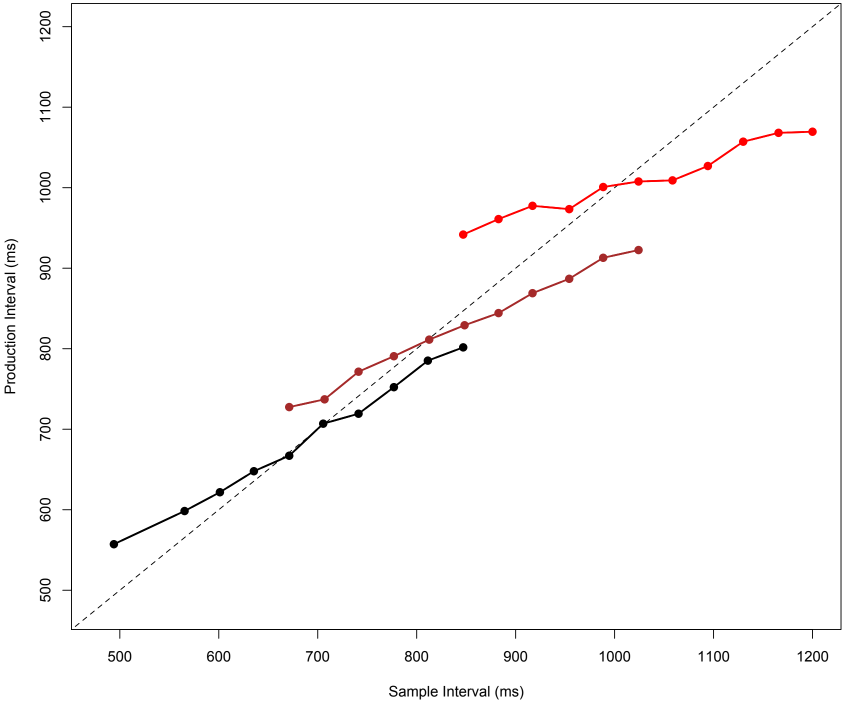
Time Estimation Formalized



Default "ACT-R" timing model

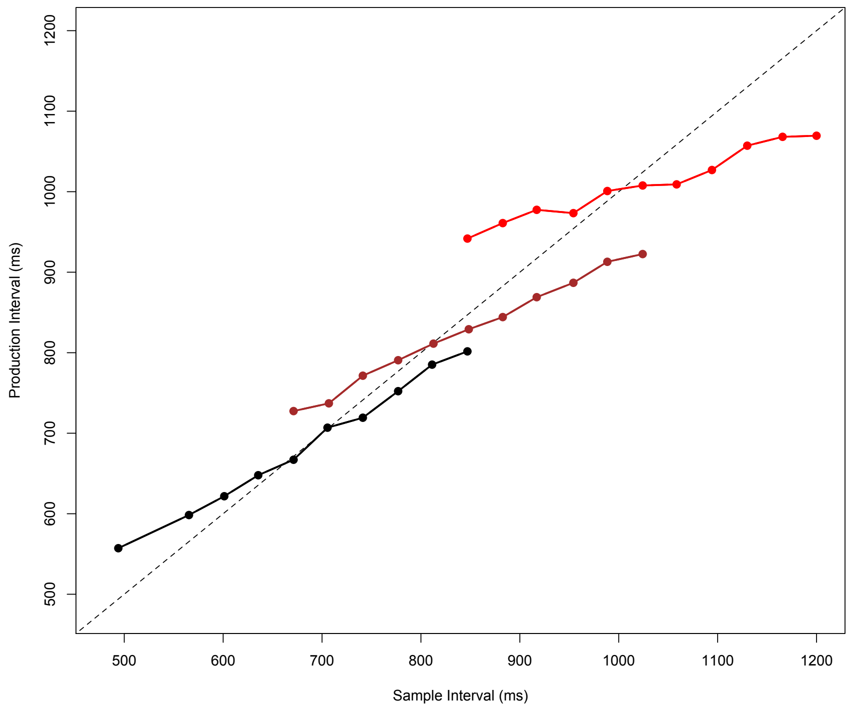
Rationale of Jazayeri & Shadlen

- *"Longer intervals engender more uncertainty"*
- More uncertain measurements result in *"increased reliance on prior expectation"*

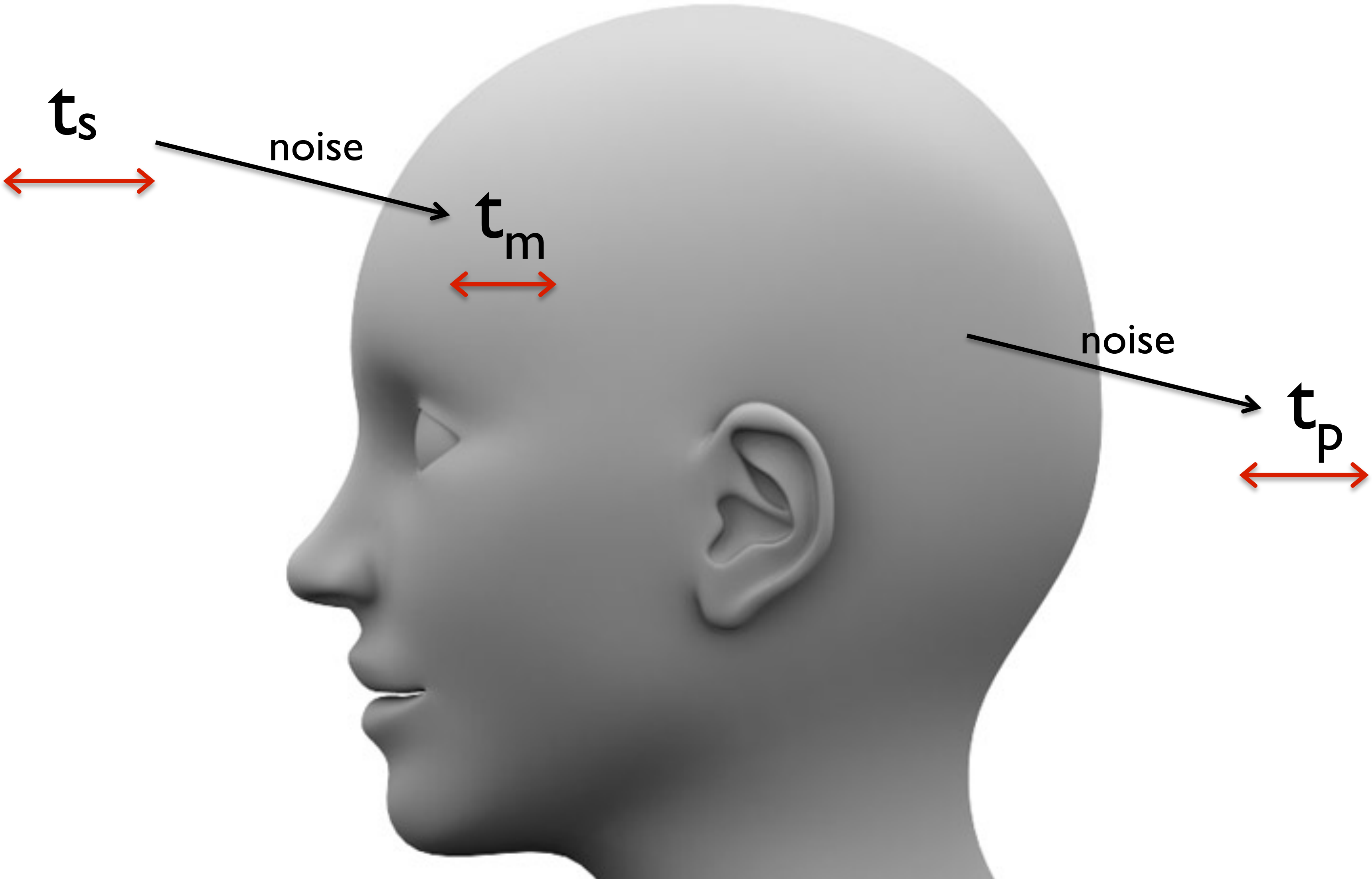


Rationale of Jazayeri & Shadlen

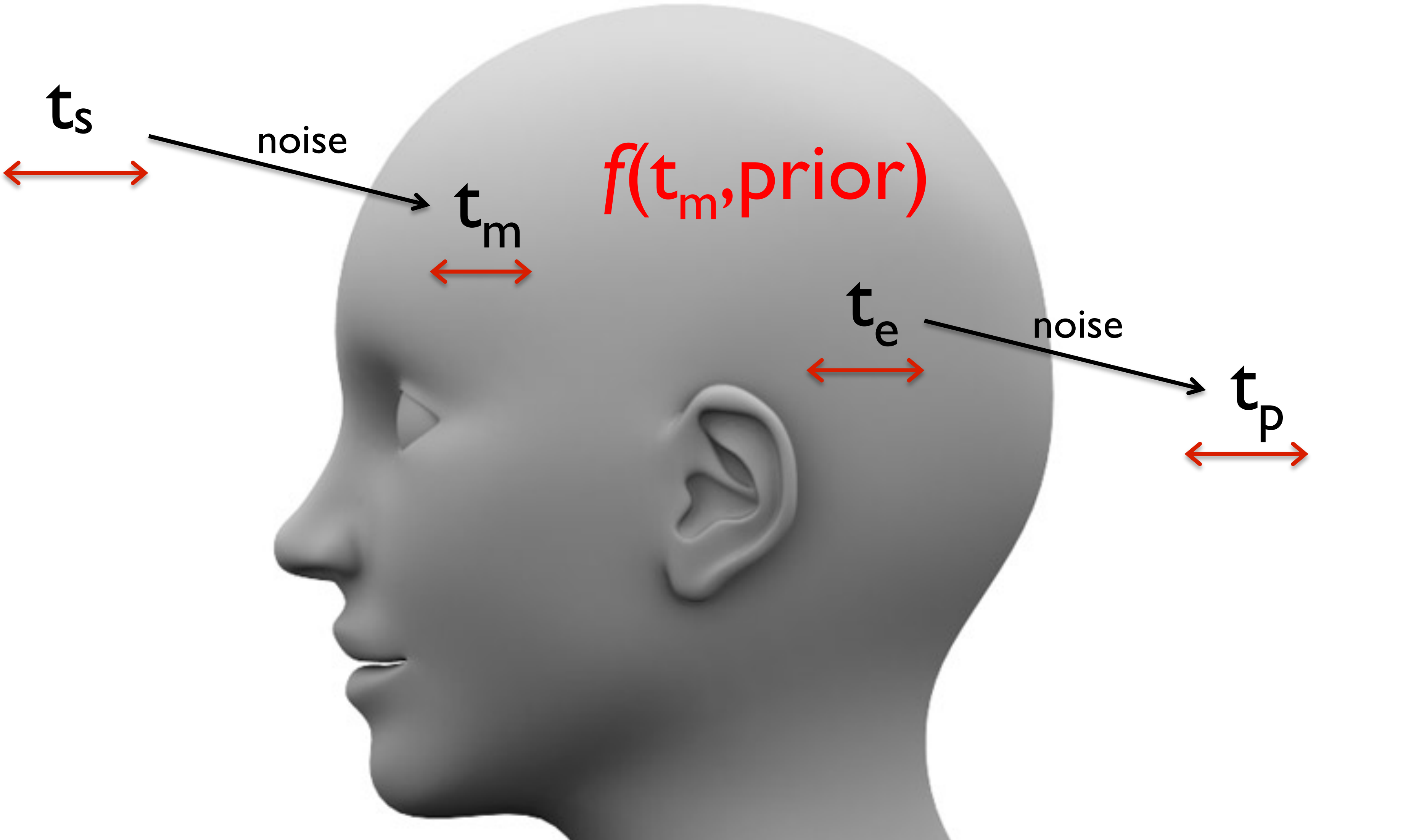
- *"Longer intervals engender more uncertainty"*
- More uncertain measurements result in *"increased reliance on prior expectation"*



From Stimulus to Reproduction



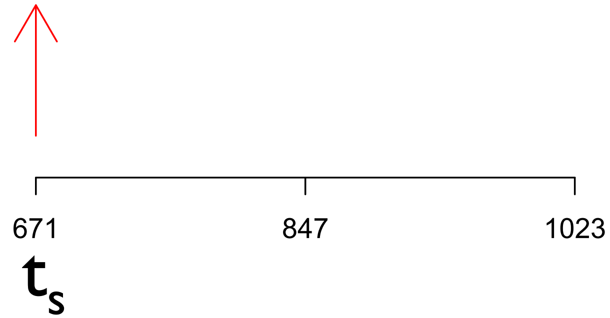
From Stimulus to Reproduction

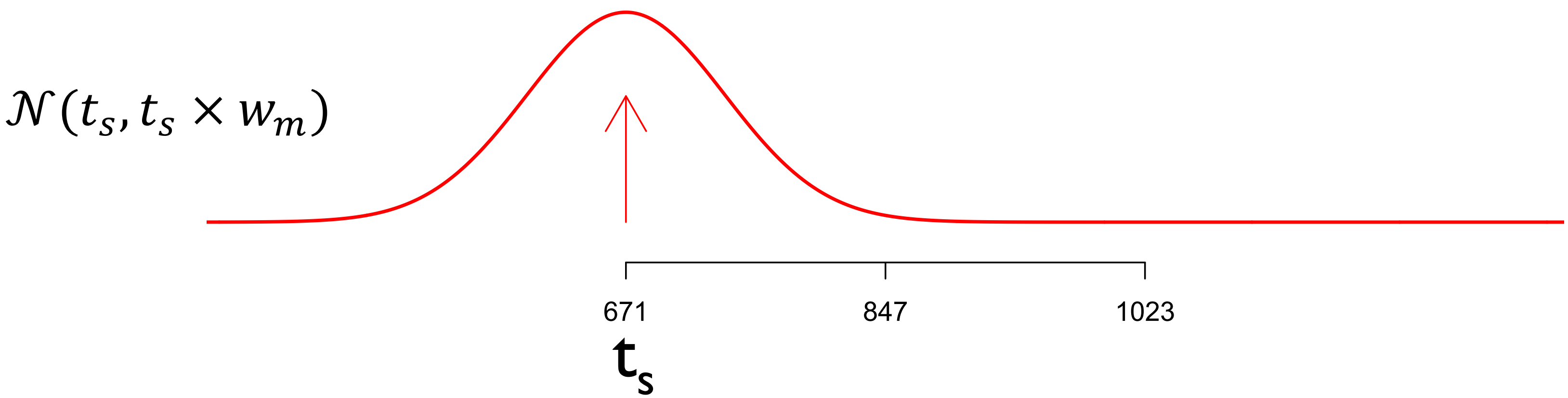


$$f(t_m, \mathrm{prior})$$

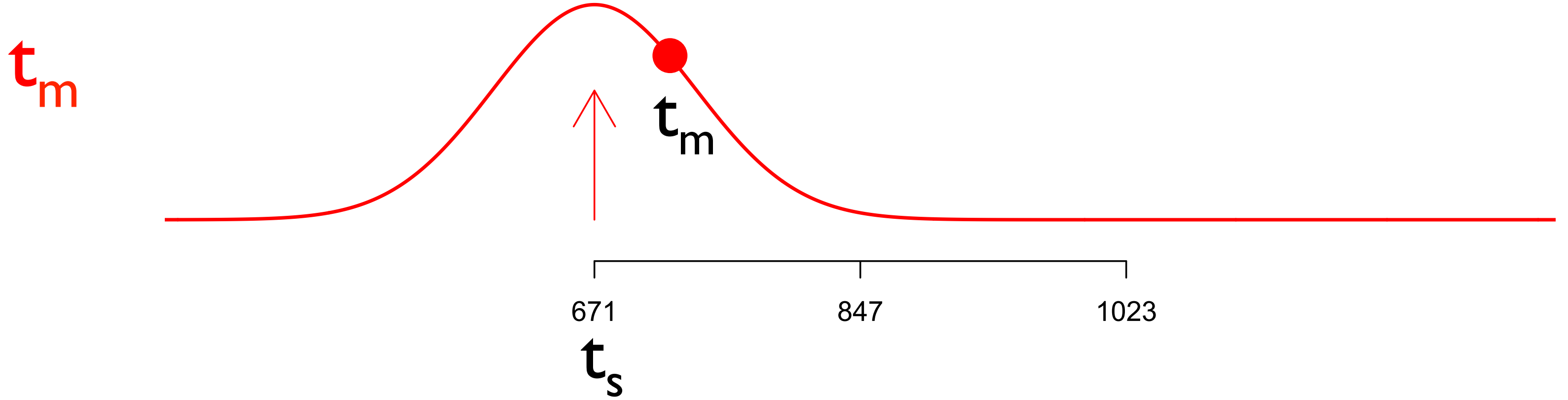
Current observation

Memory of previous durations

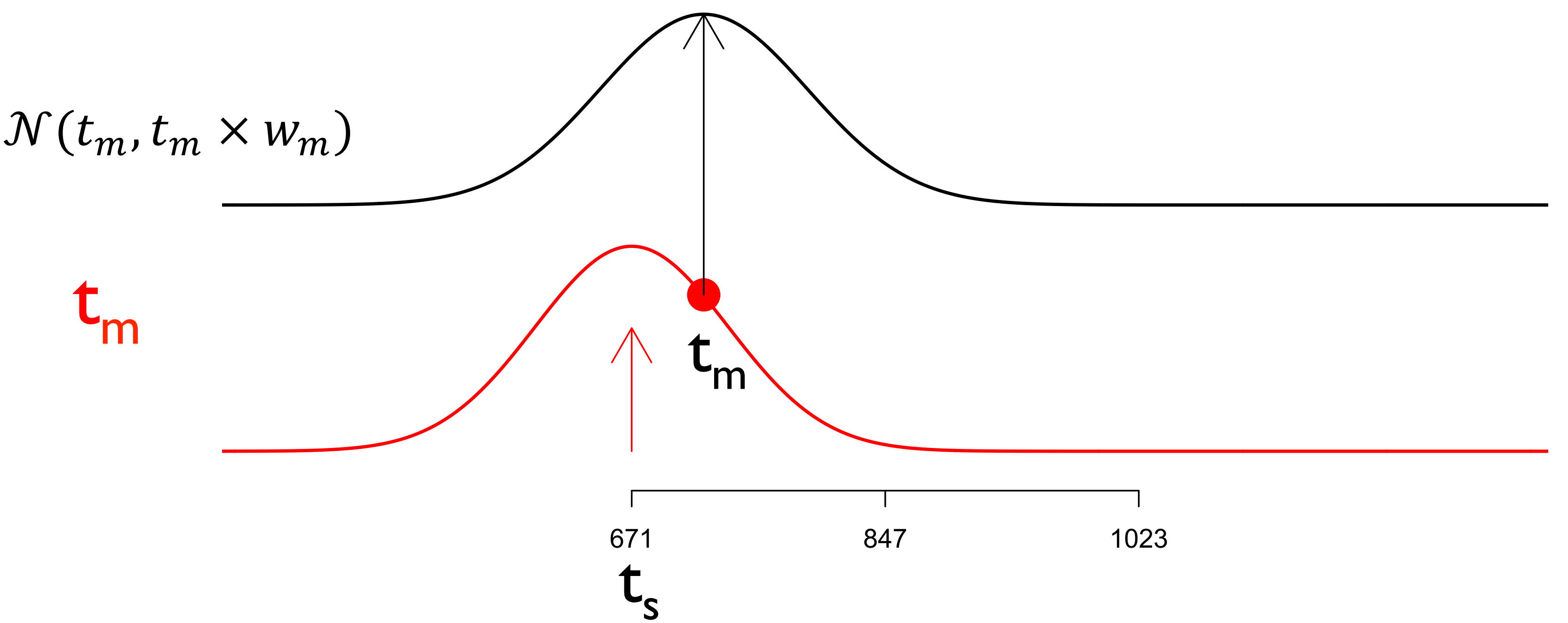

671
847
1023
t_s



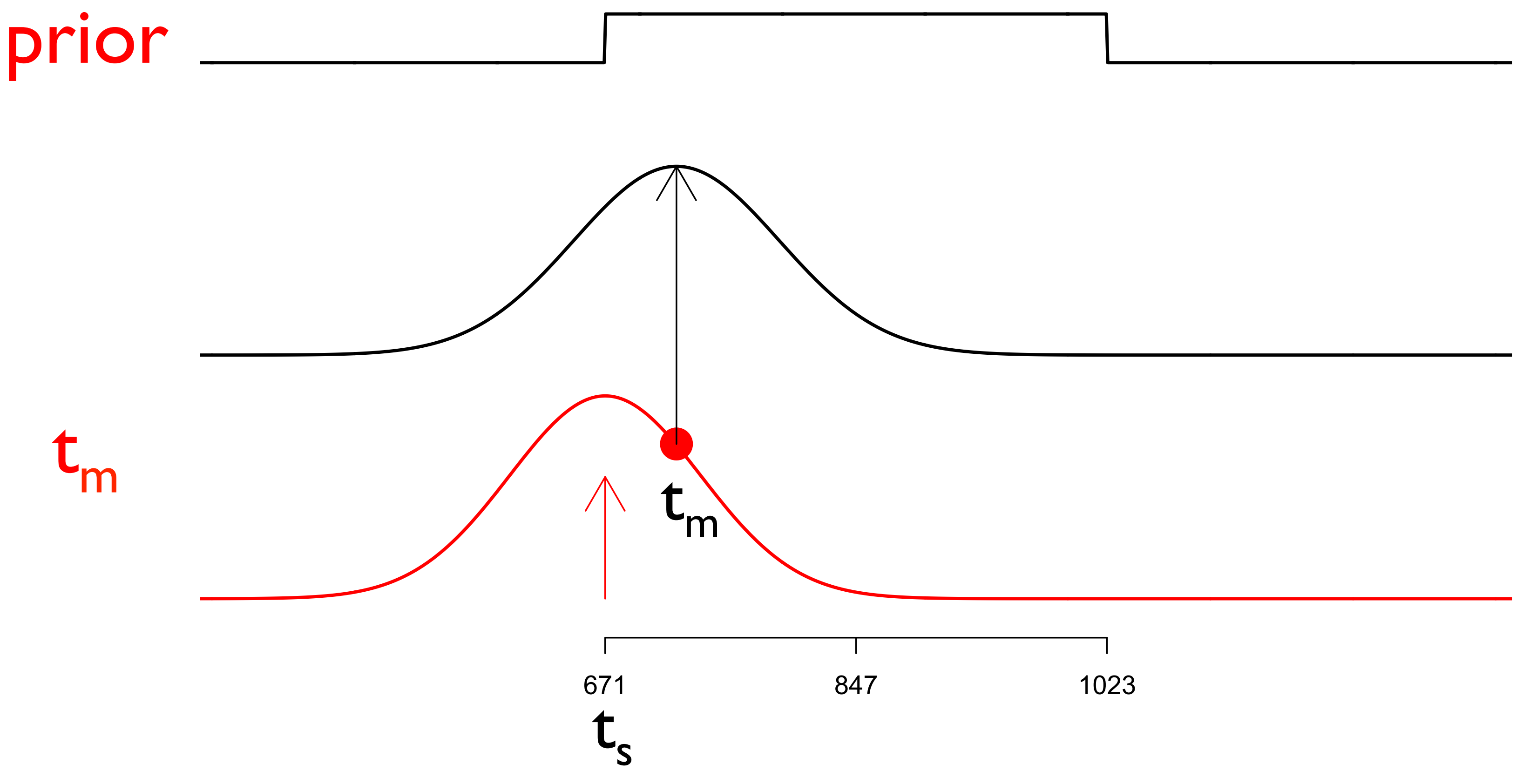
$\mathcal{N}(t_s, t_s \times w_m)$
671
847
1023
t_s



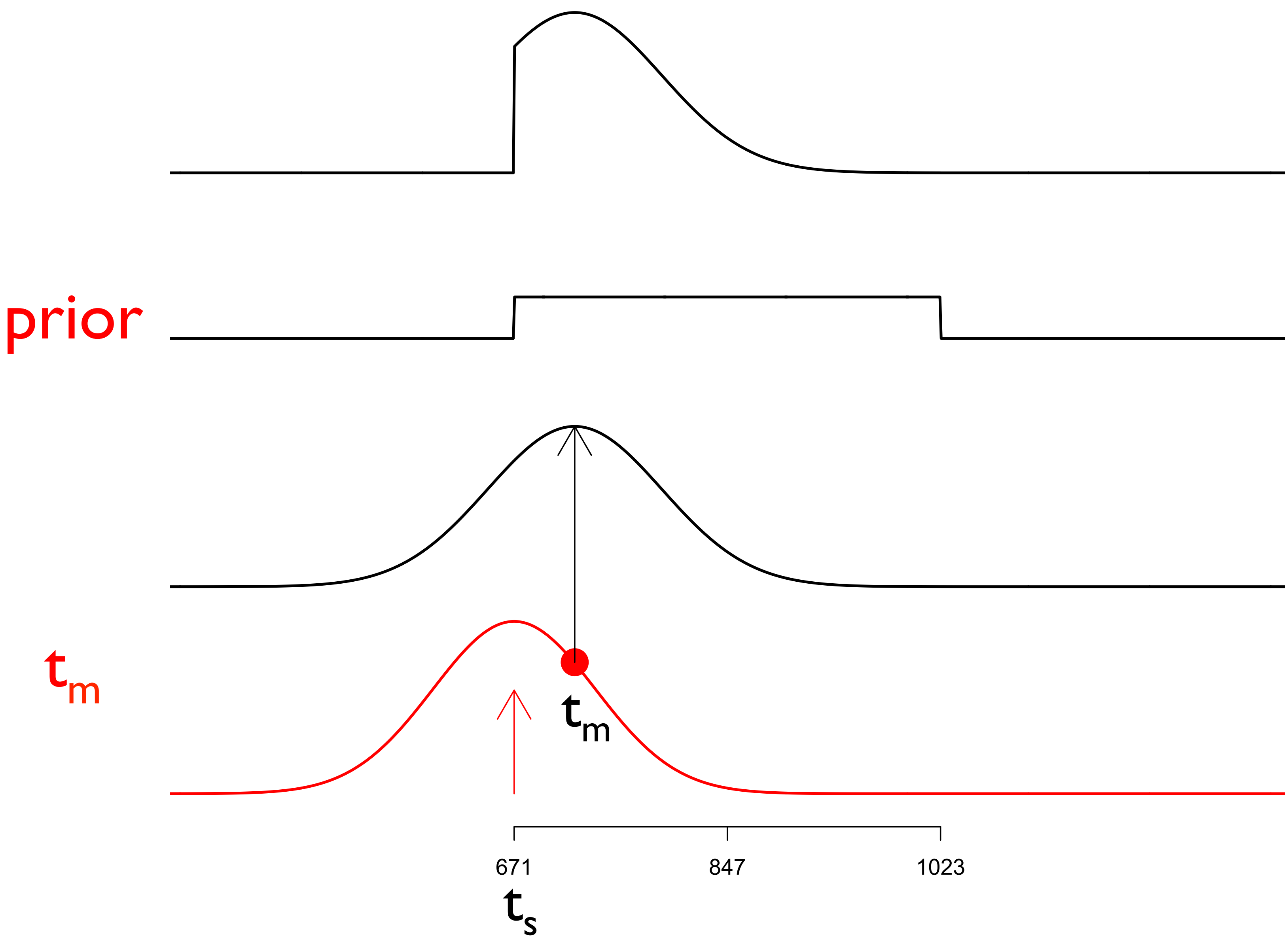
t_m
t_m
671
847
1023
t_s



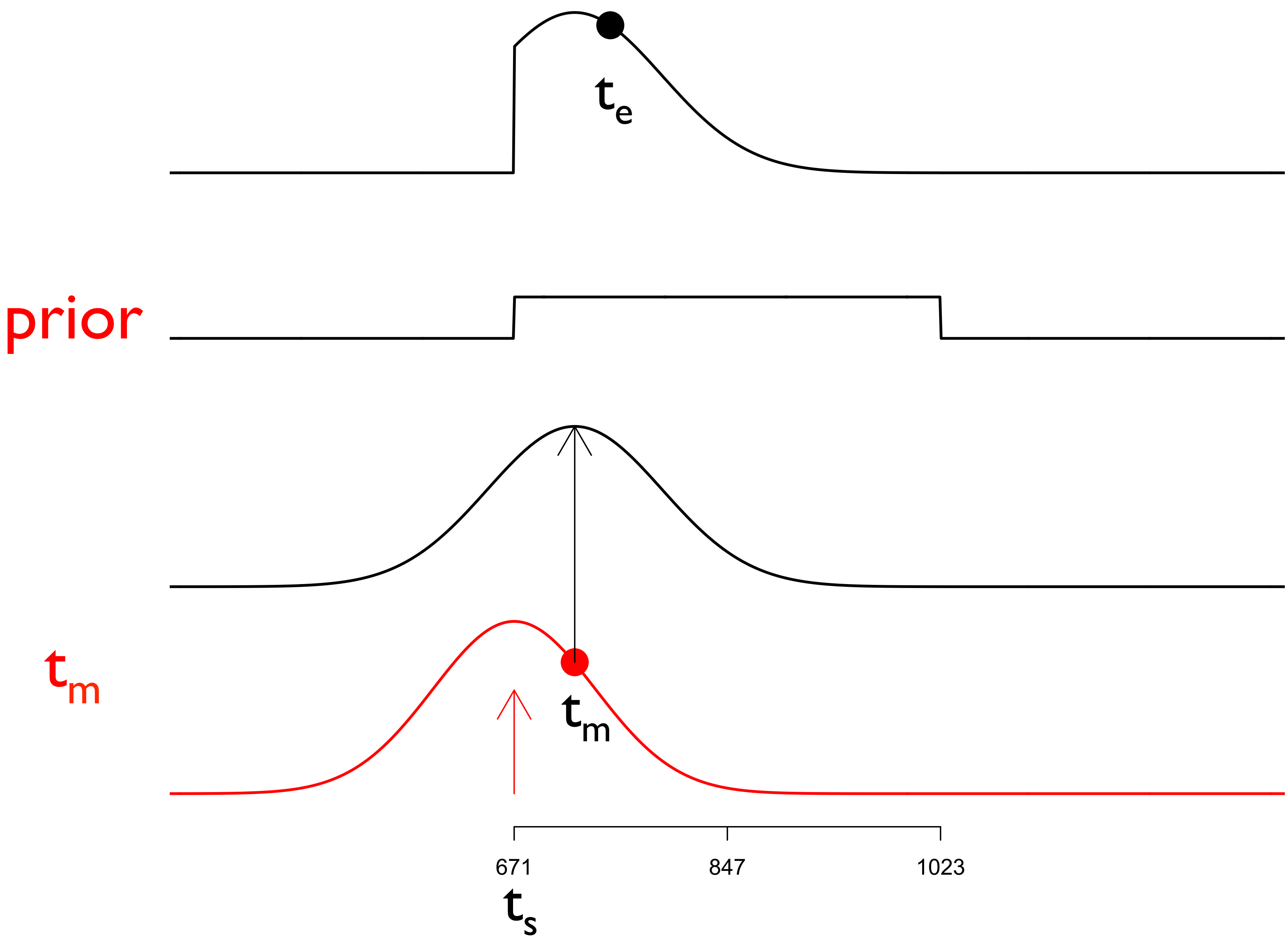
$\mathcal{N}(t_m, t_m \times w_m)$
t_m
t_m
671
847
1023
t_s



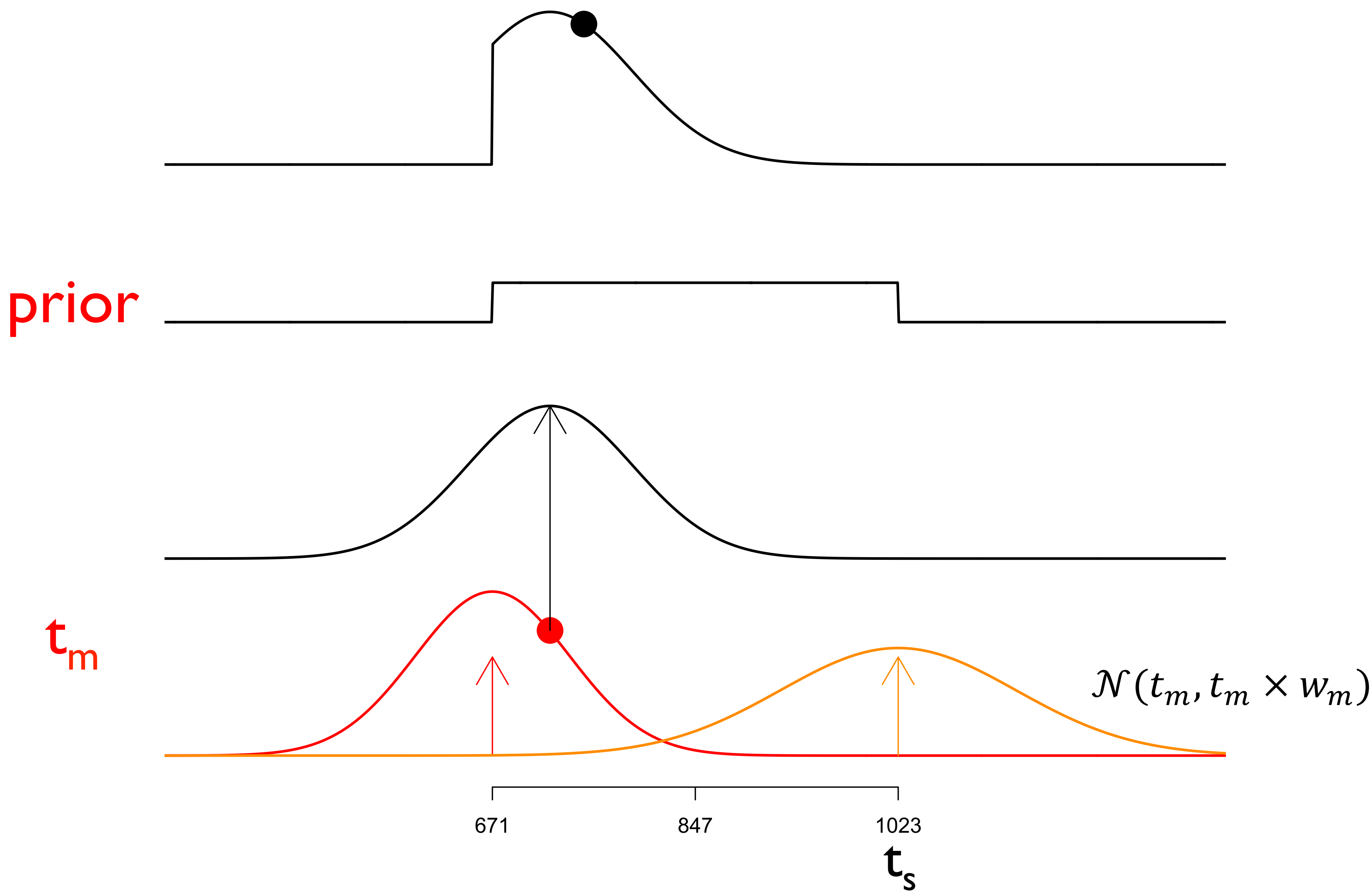
prior
t_m
t_m
671
847
1023
t_s

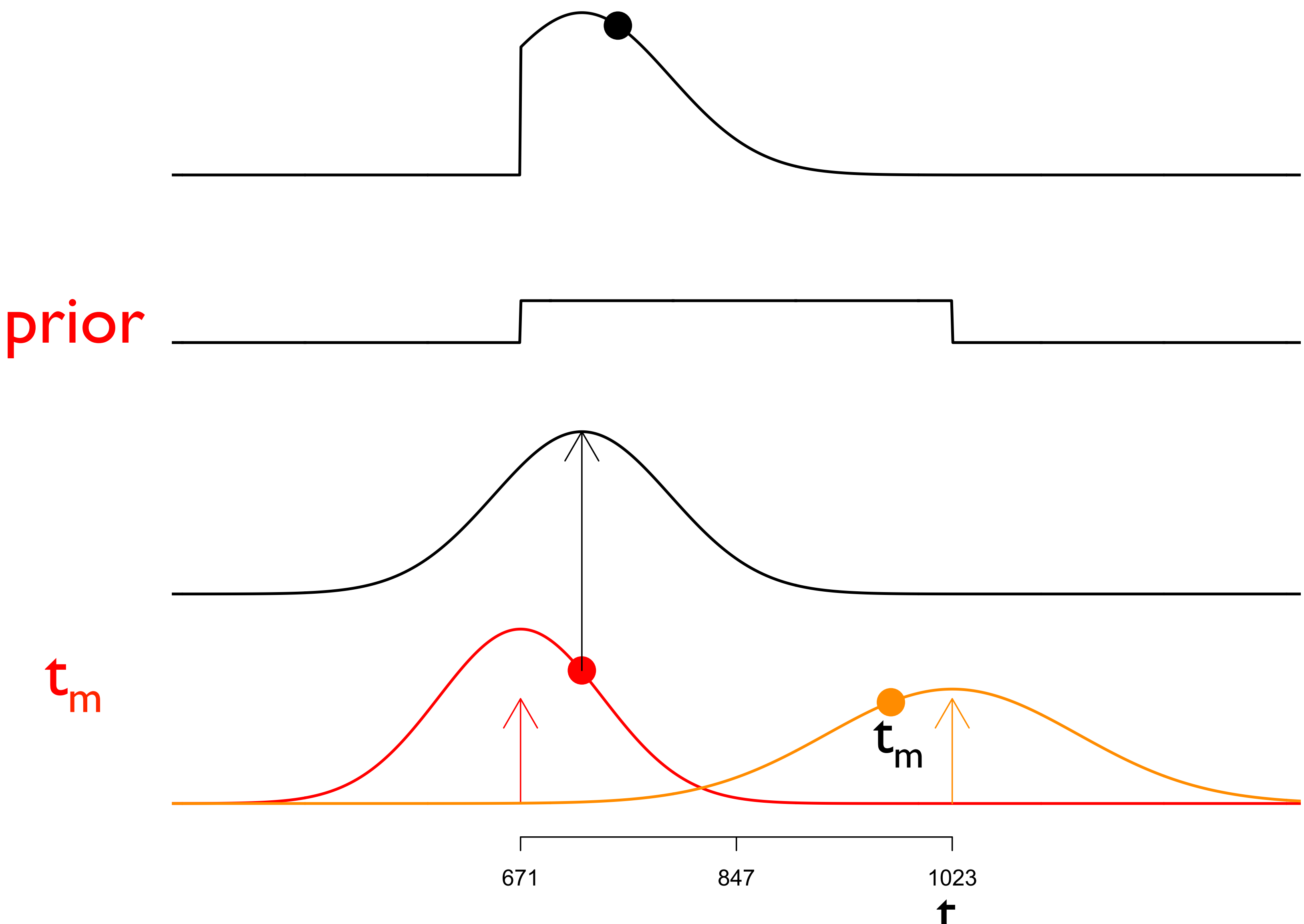


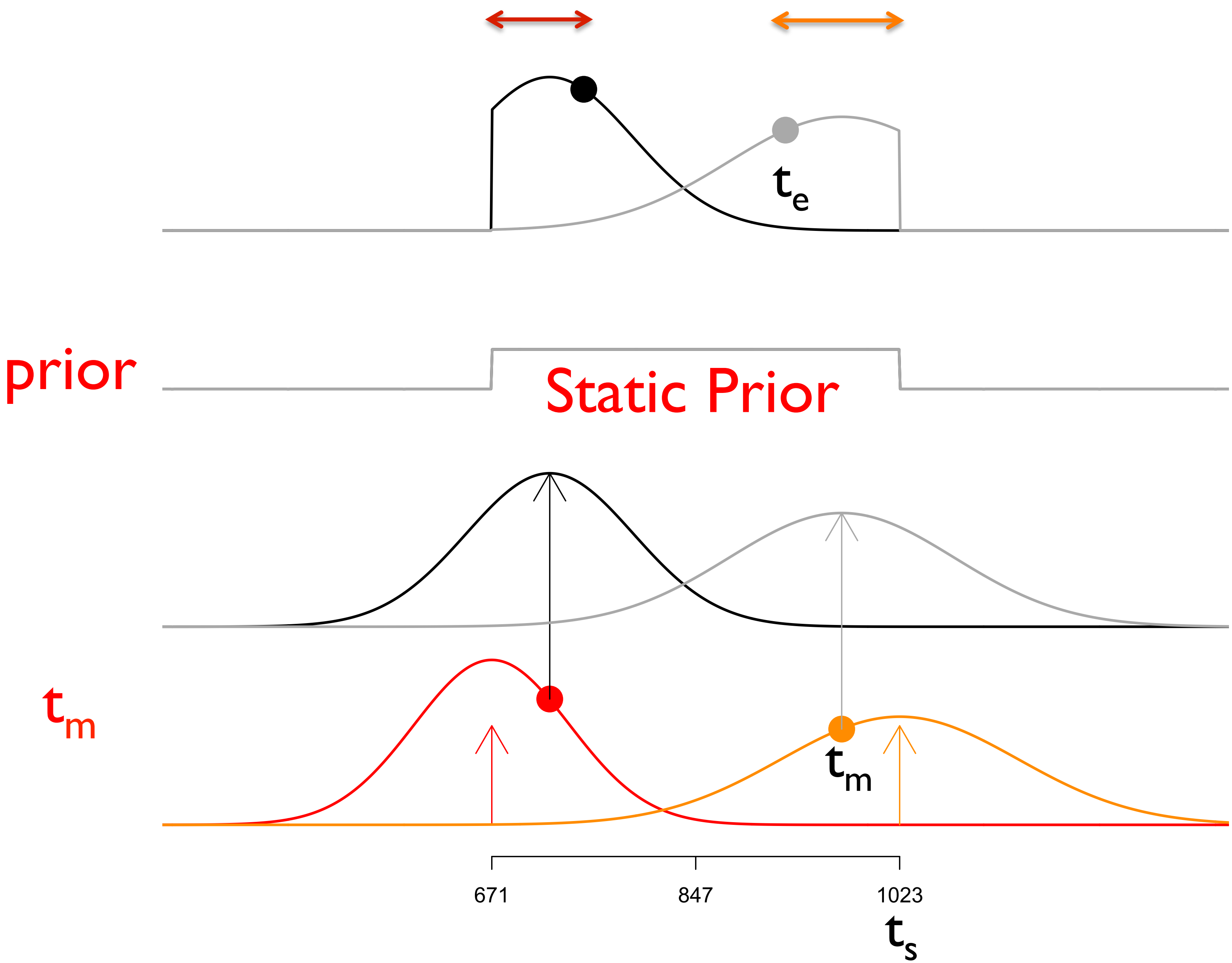
prior
t_m
t_m
671
847
1023
t_s



t_e
prior
t_m
t_m
671
847
1023
t_s

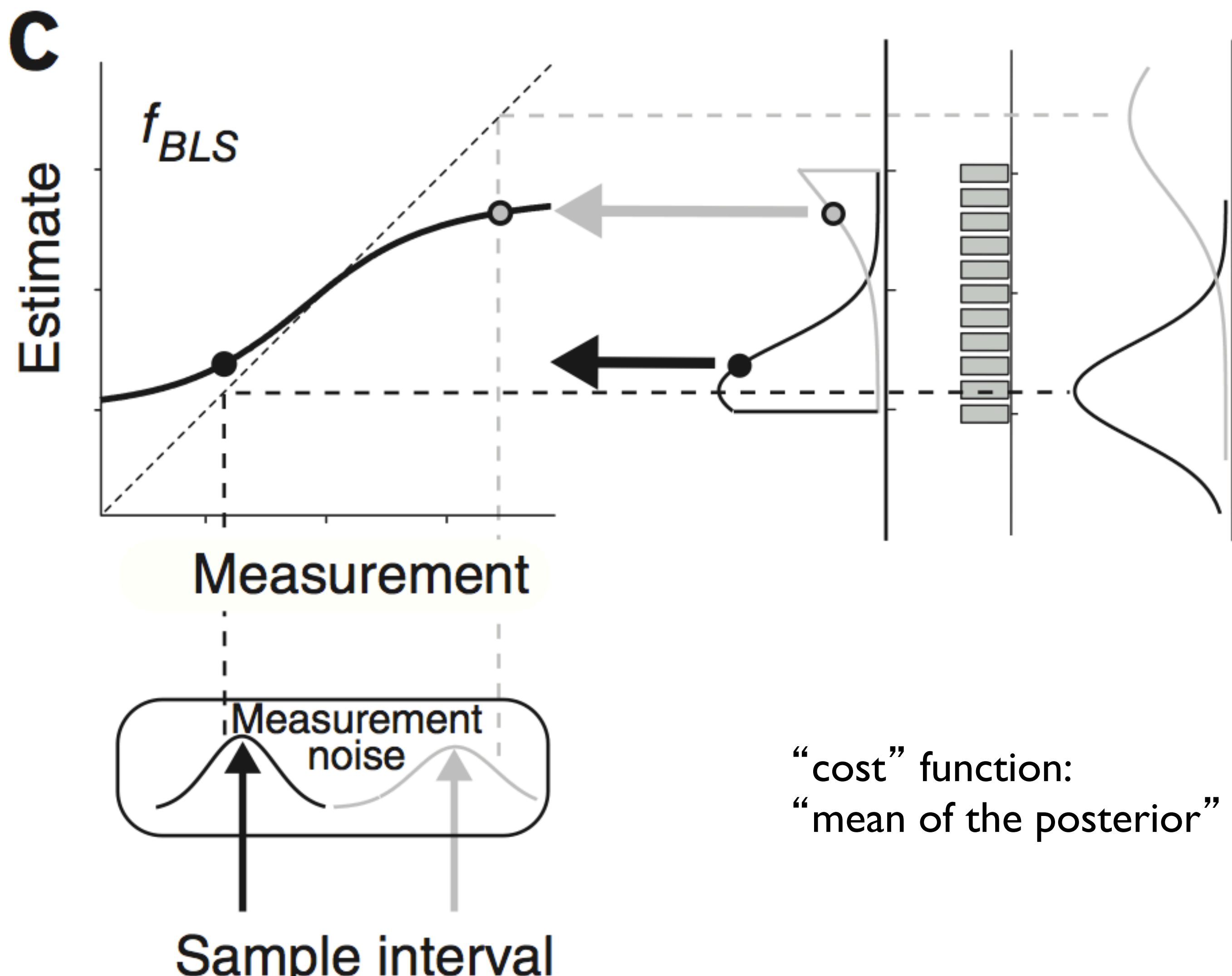

prior
t m
$\mathcal{N}(t_m, t_m \times w_m)$
671
847
1023
t s


prior
t_m
t_m
671
847
1023
t_s



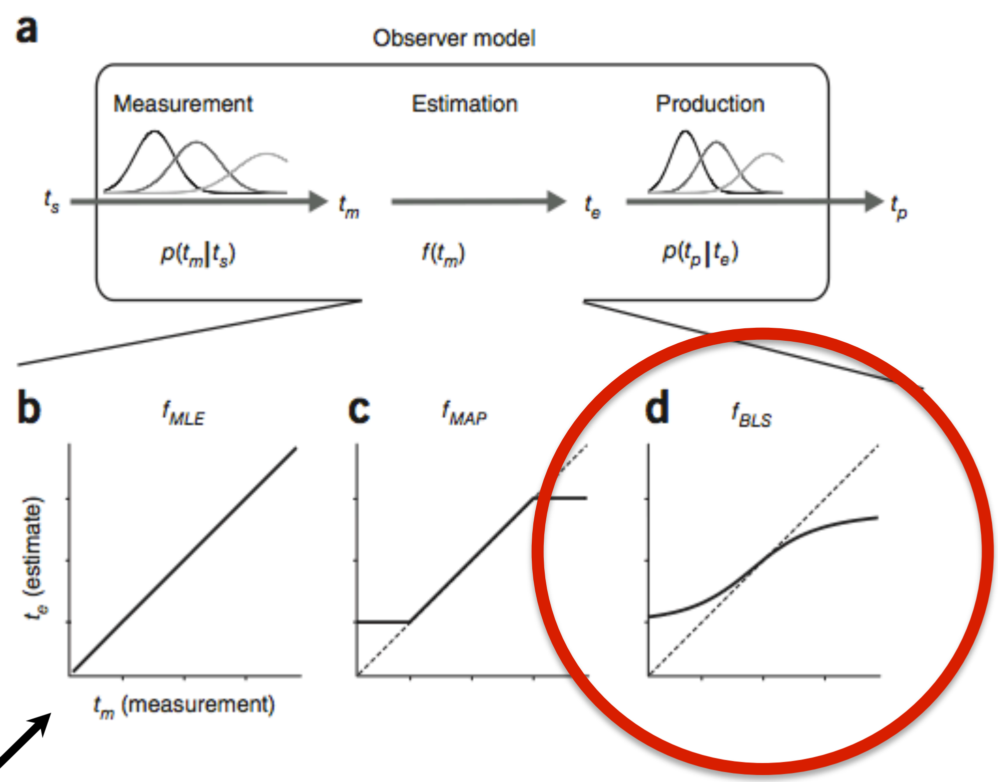
prior
Static Prior
t_e
t_m
t_m
671
847
1023
t_s

Bayes least-squares
Weighted Context



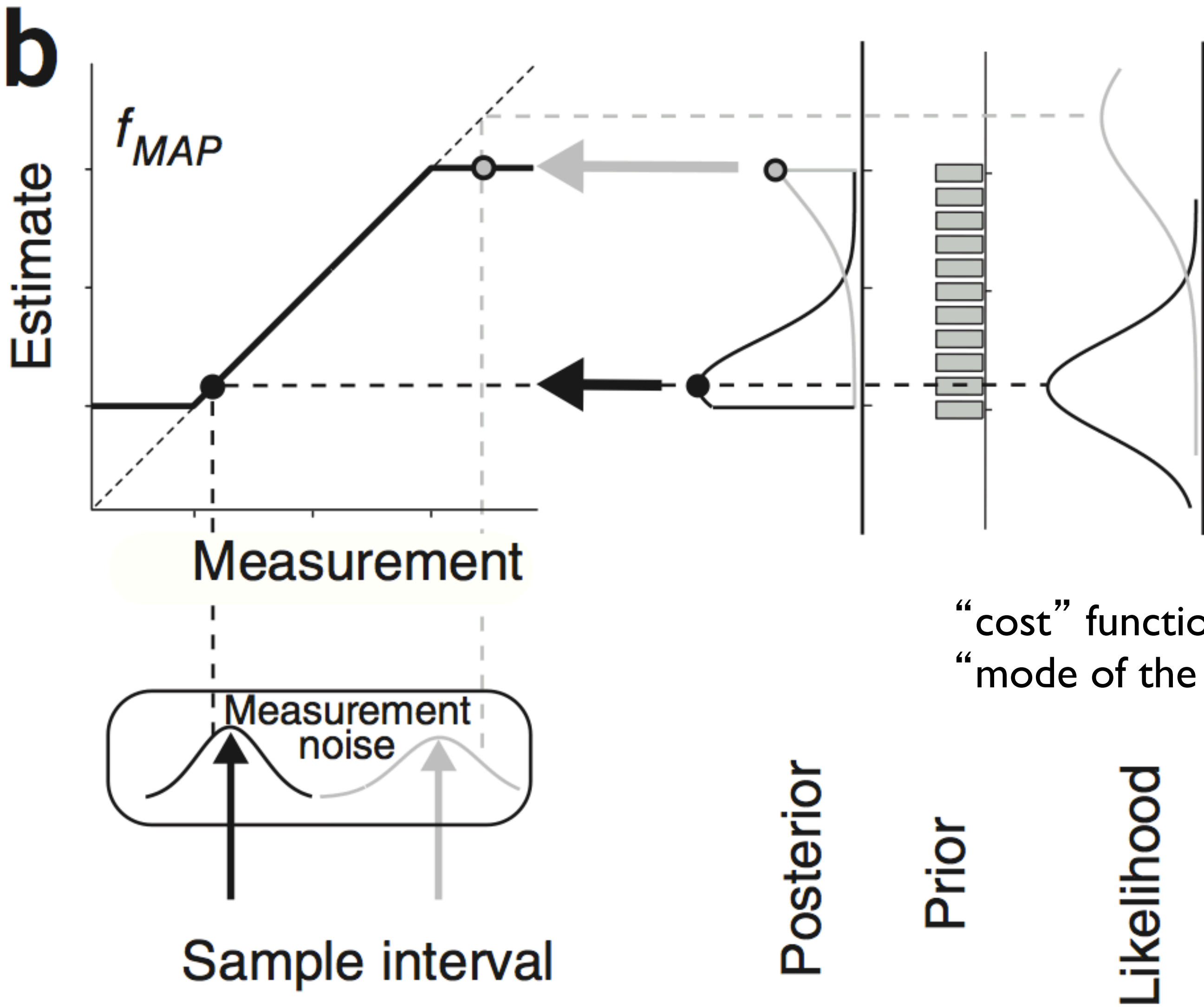
“cost” function:
“mean of the posterior”

Time Estimation Formalized



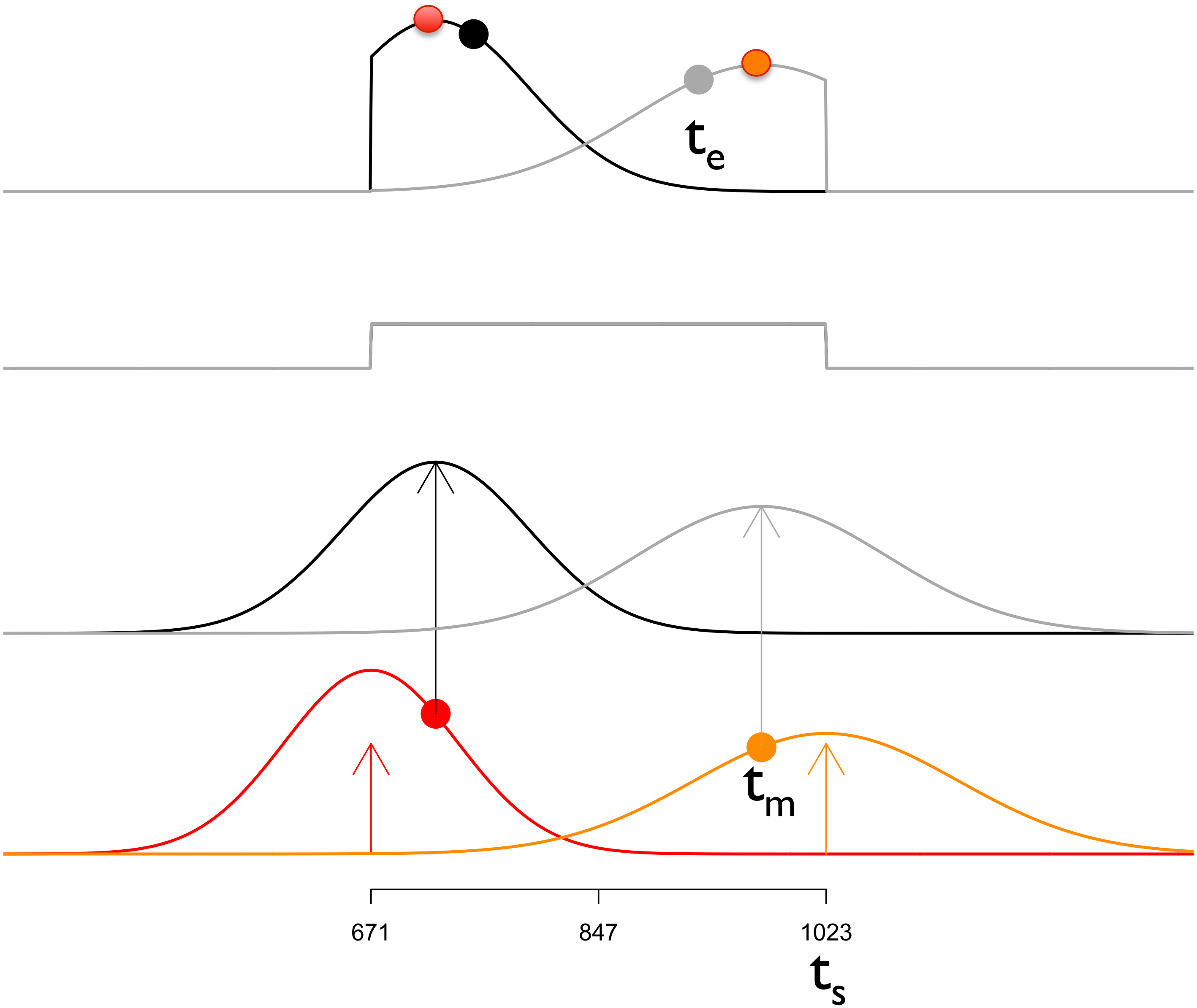
Default "ACT-R" timing model

Maximum a posteriori
“Binary Context”



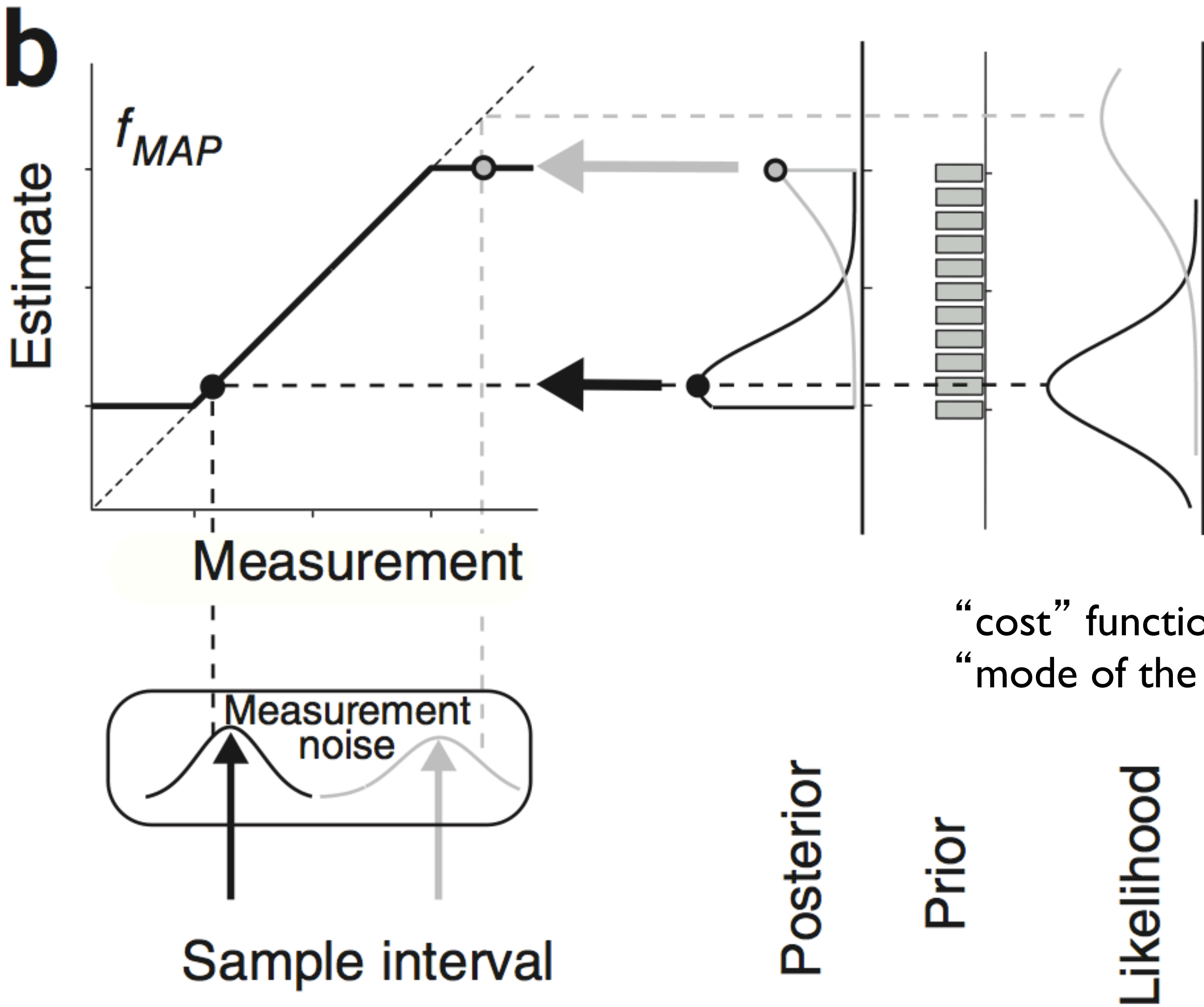
Alternatively, a Bayesian observer would combine the likelihood function and the prior and use some statistic to map the resulting posterior probability distribution onto an estimate.

“cost” function:
“mode of the posterior”



t_e
t_m
671
847
1023
t_s

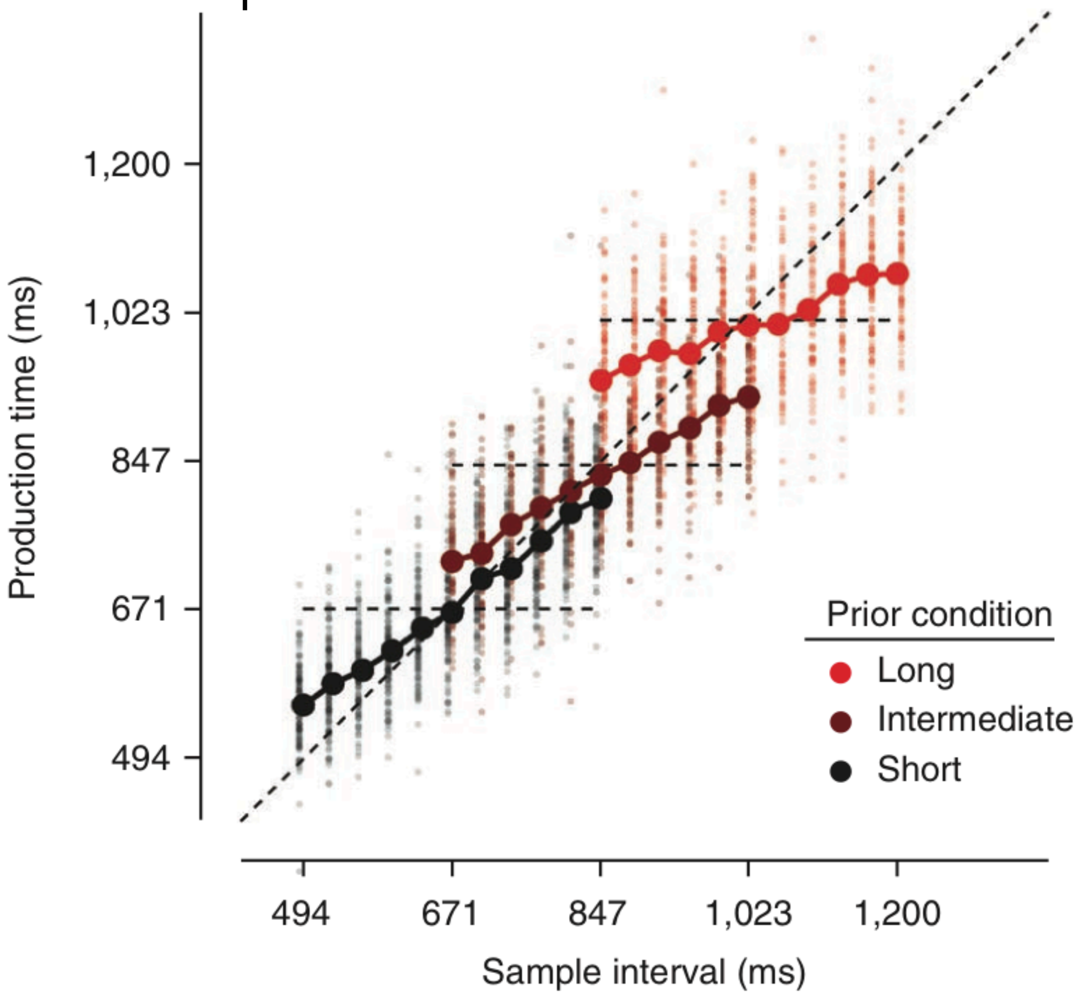
Maximum a posteriori
"Binary Context"



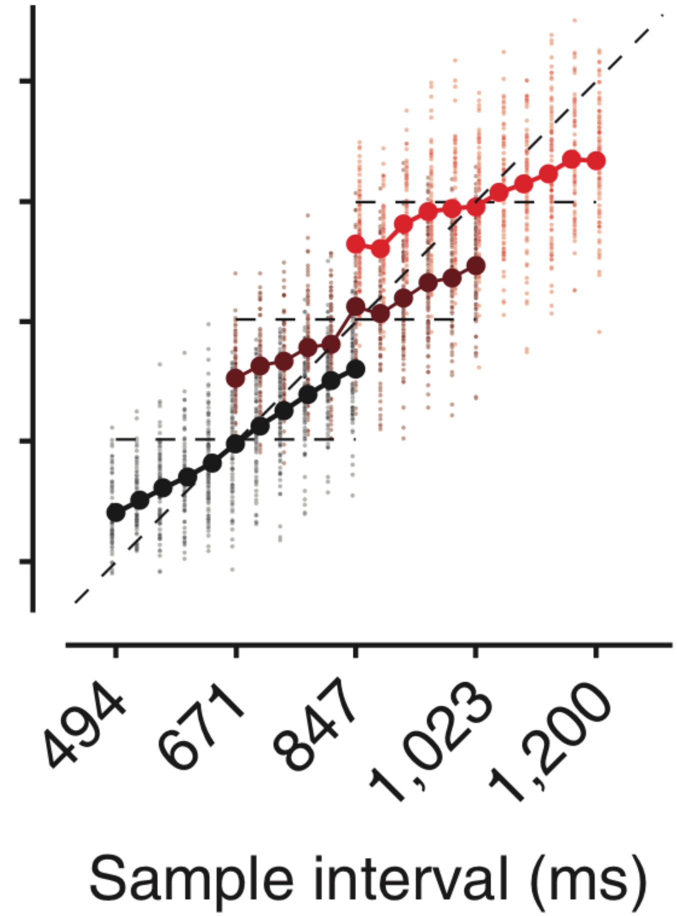
Alternatively, a Bayesian observer would combine the likelihood function and the prior and use some statistic to map the resulting posterior probability distribution onto an estimate.

"cost" function:
"mode of the posterior"

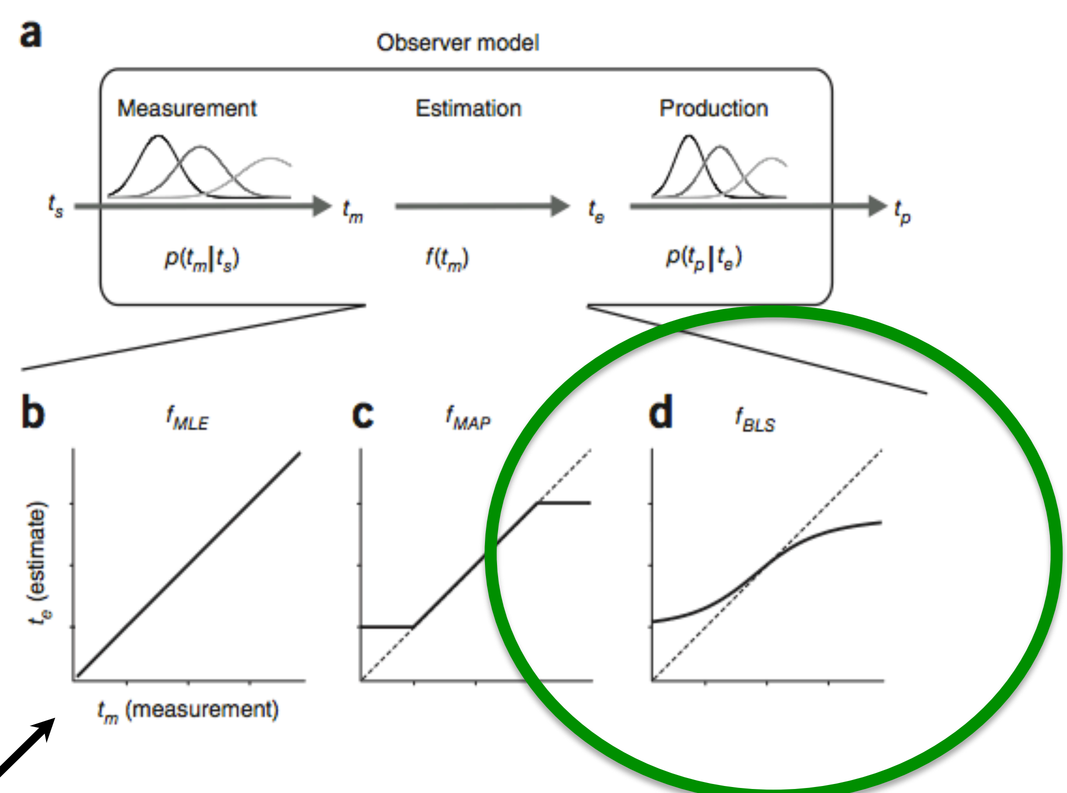
Empirical data



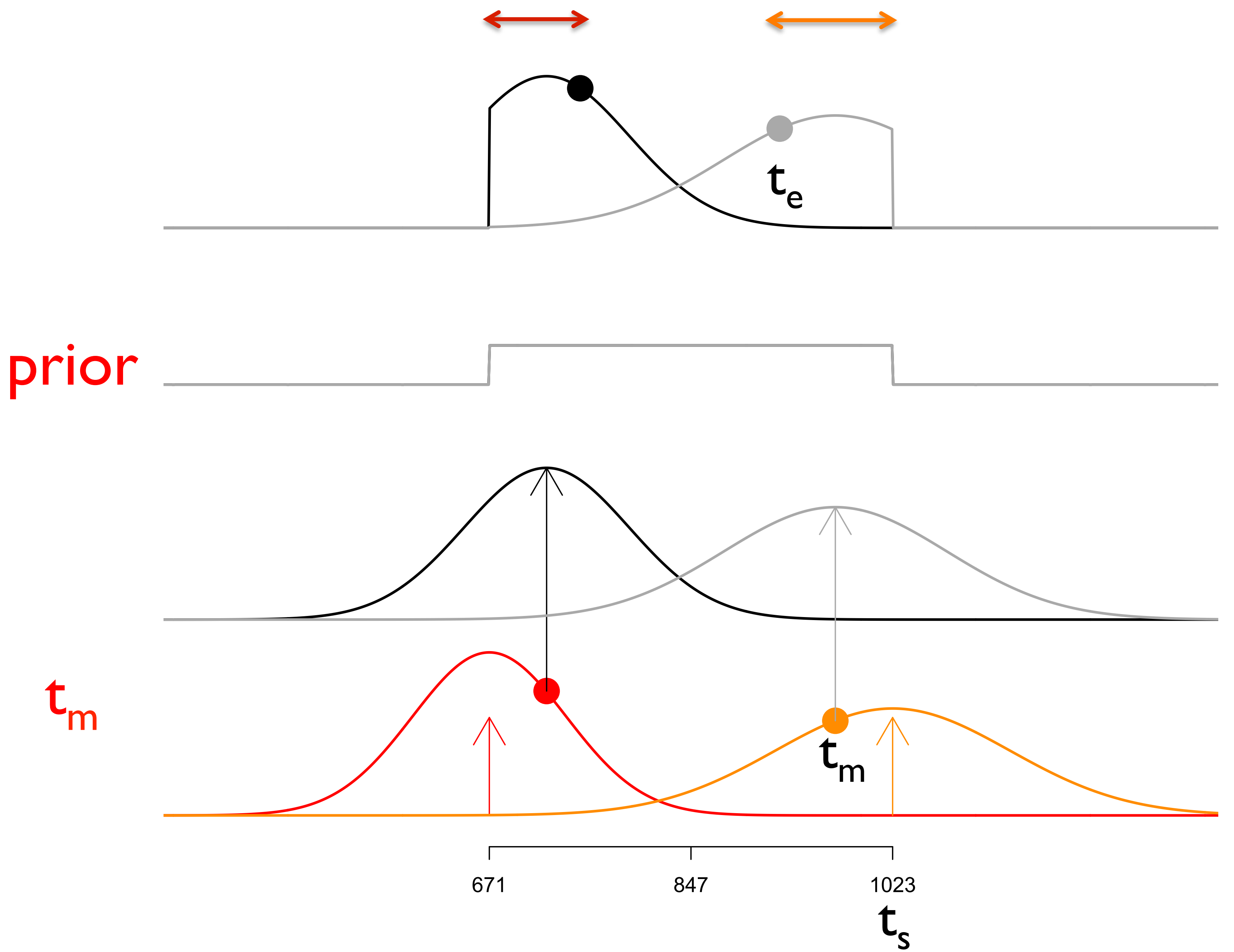
BLS Model



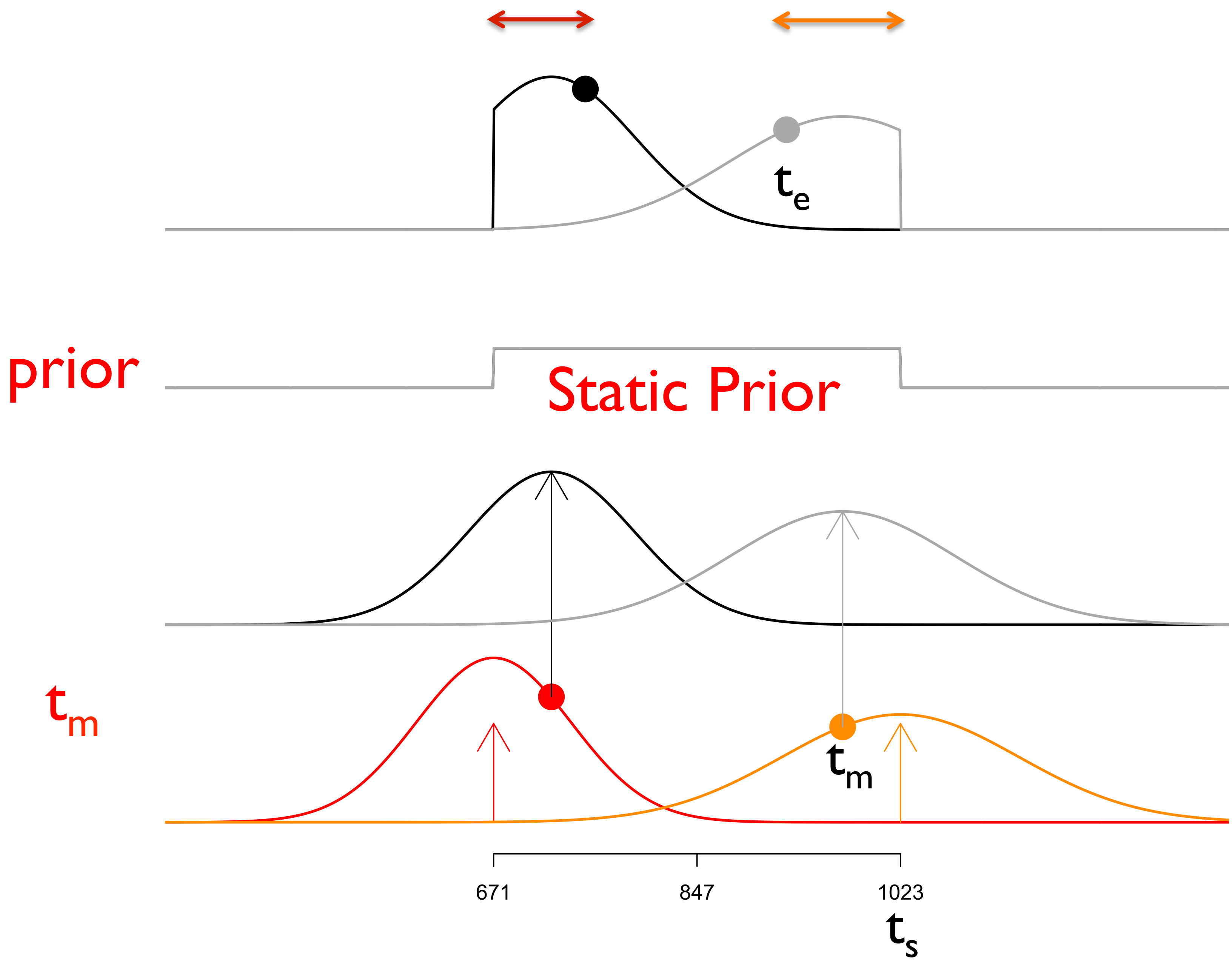
Time Estimation Formalized



Default "ACT-R" timing model

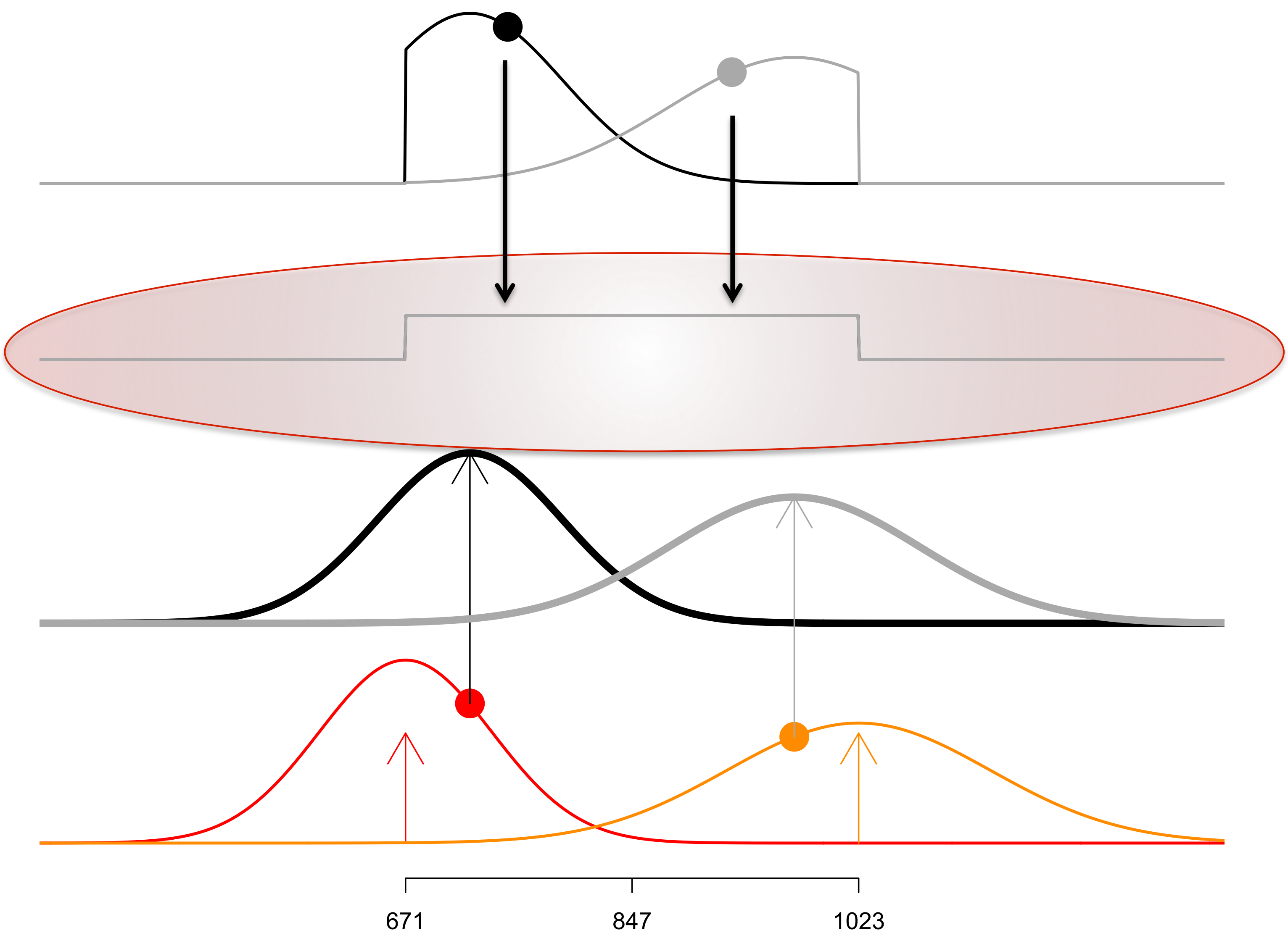


t_e
prior
t_m
t_m
671
847
1023
t_s



t_e
prior
Static Prior
t_m
t_m
671
847
1023
t_s

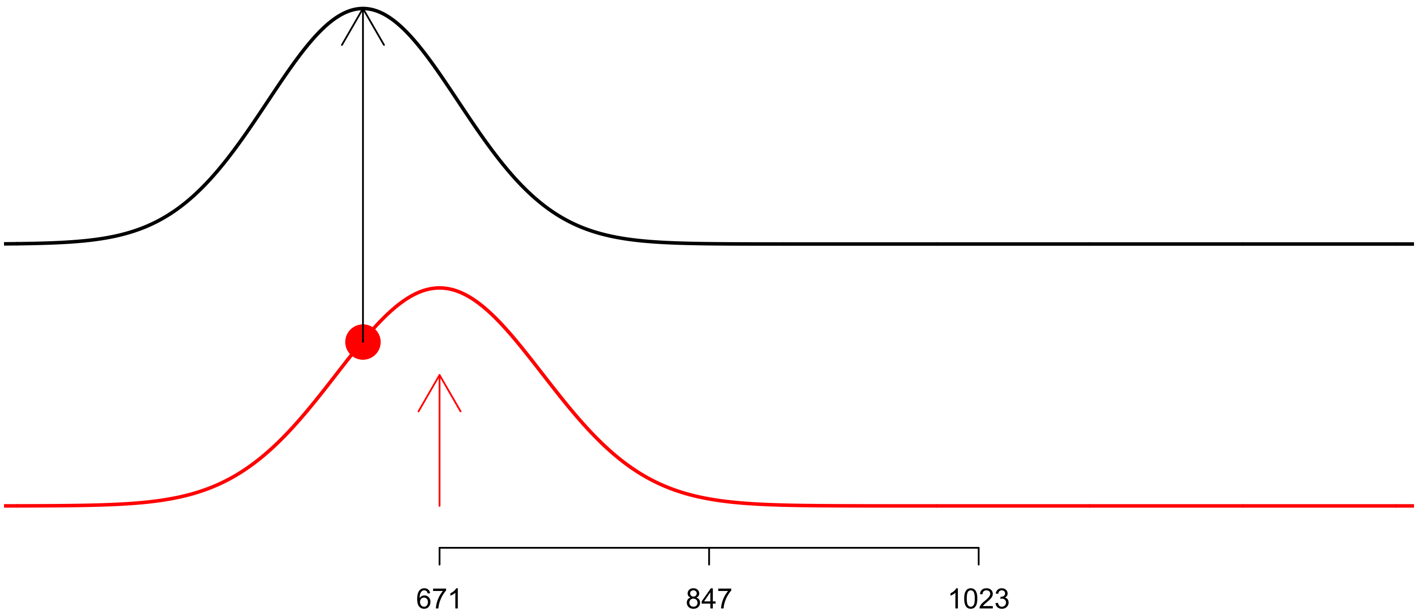
However…

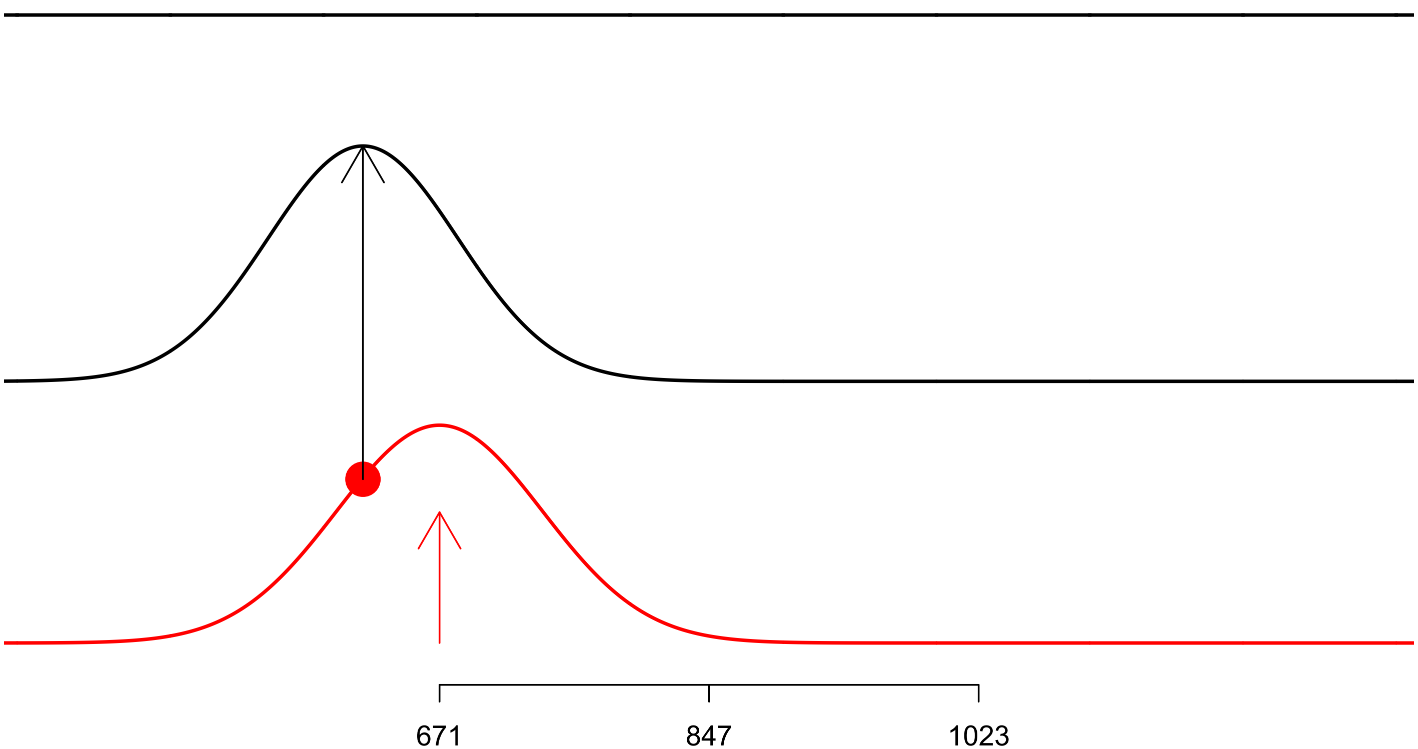

671
847
1023

Jones & Love (2011, BBS)

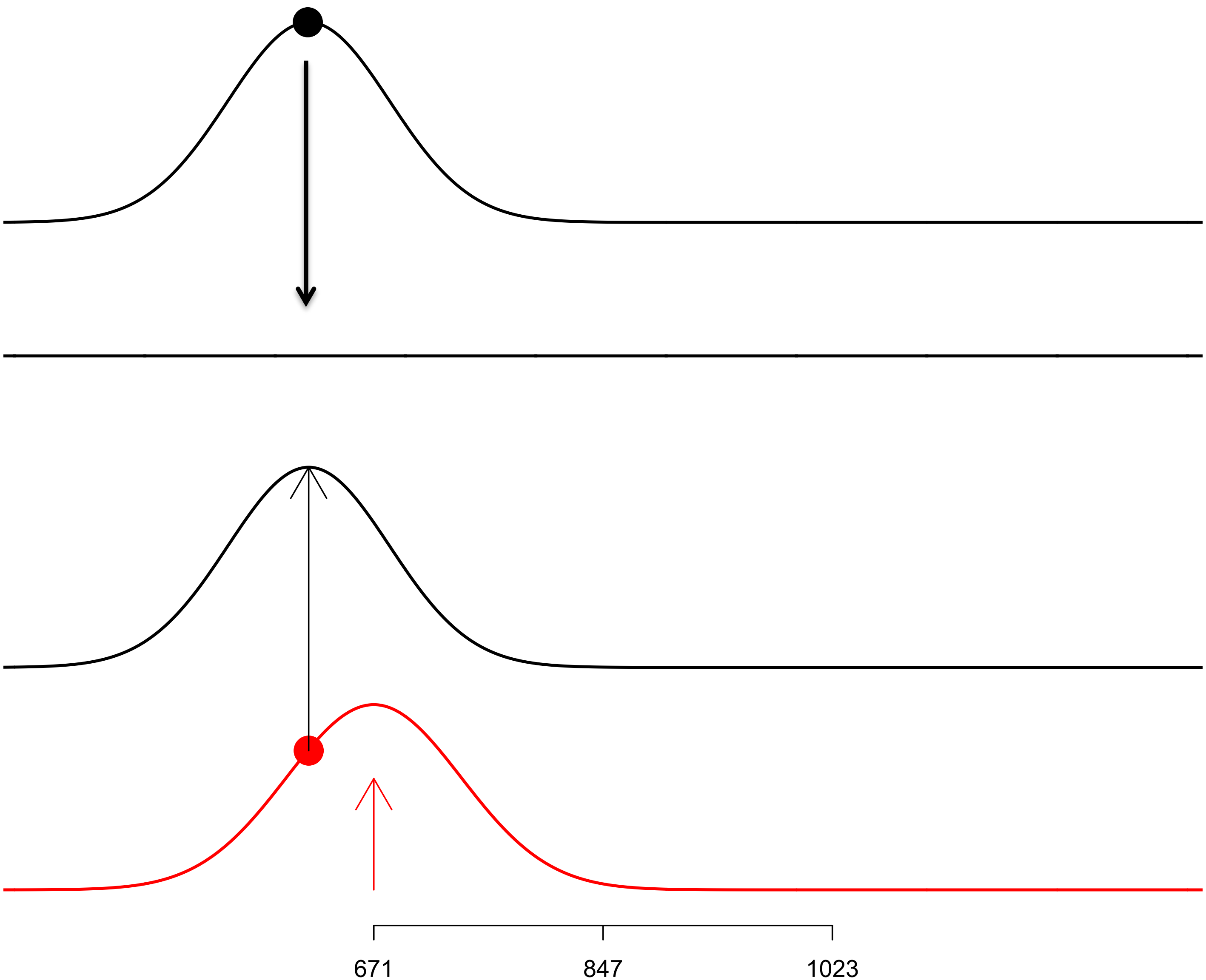
“[T]he prior can be a strong point of the model if it is derived from empirical statistics of real environments. However, more commonly **the prior is chosen ad hoc**, providing substantial unconstrained flexibility to models that are advocated as rational and assumption-free.”

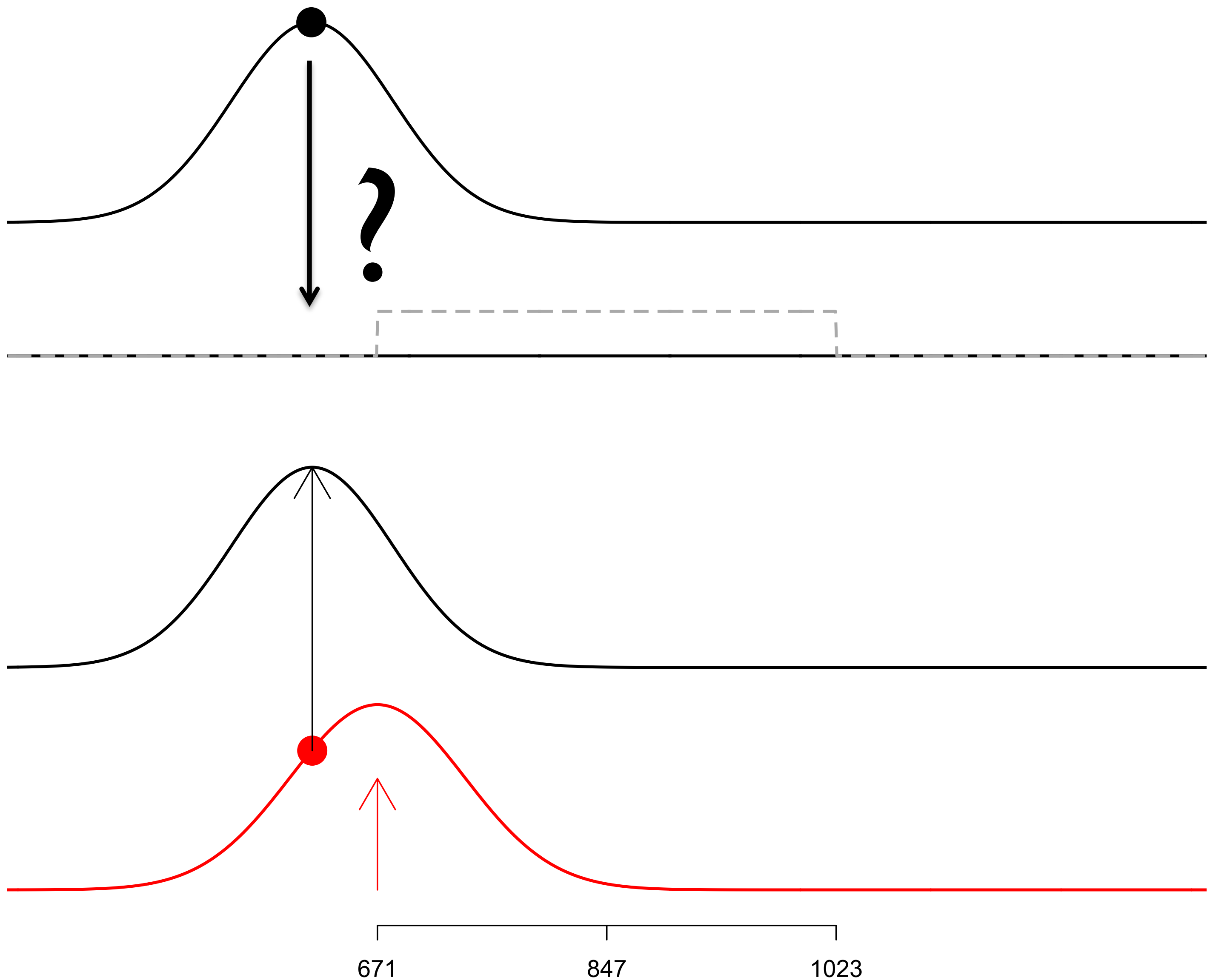
(p173-4)


671
847
1023

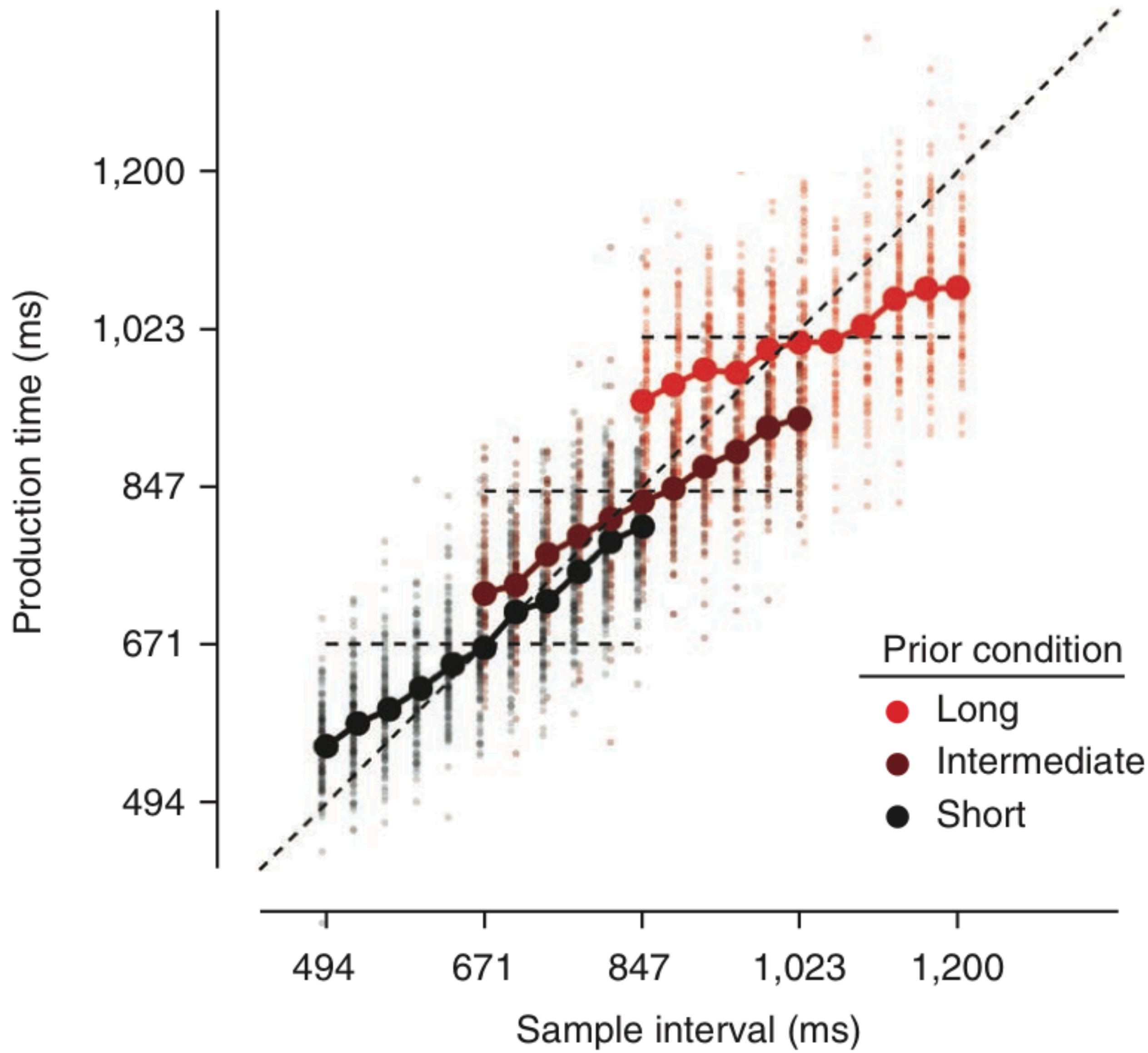


671
847
1023


671
847
1023

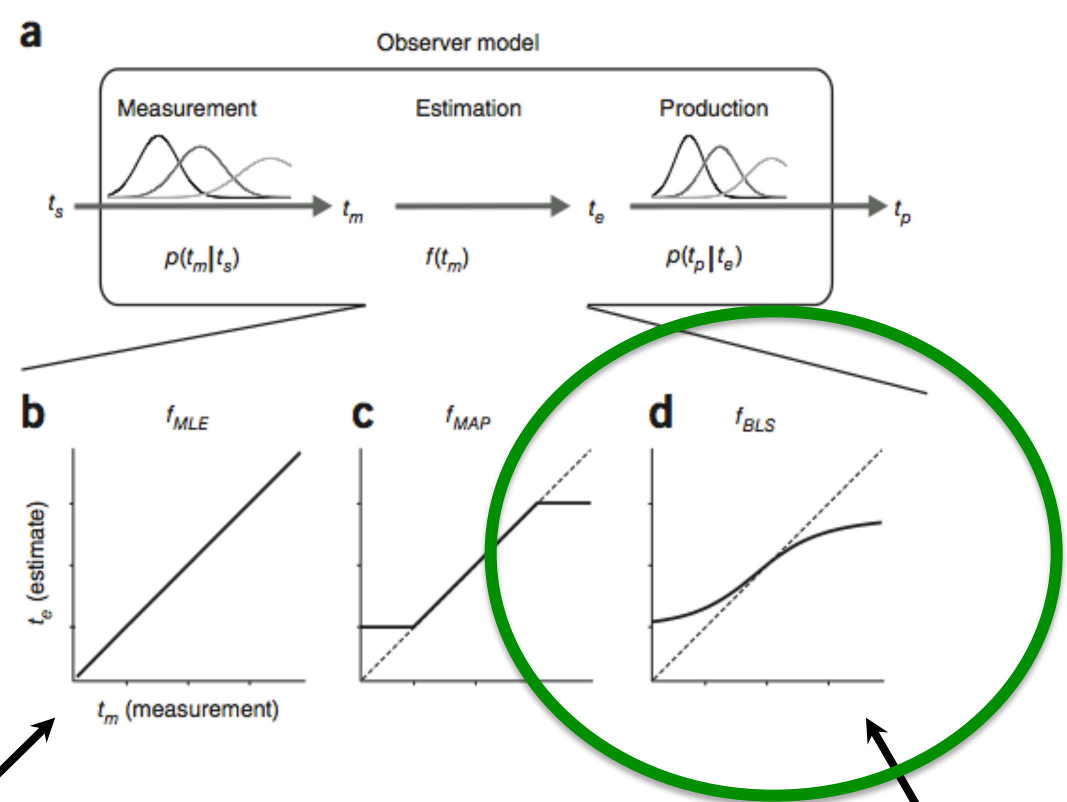

?
671
847
1023

Empirical data



And what about the dynamics of memory?

Time Estimation Formalized

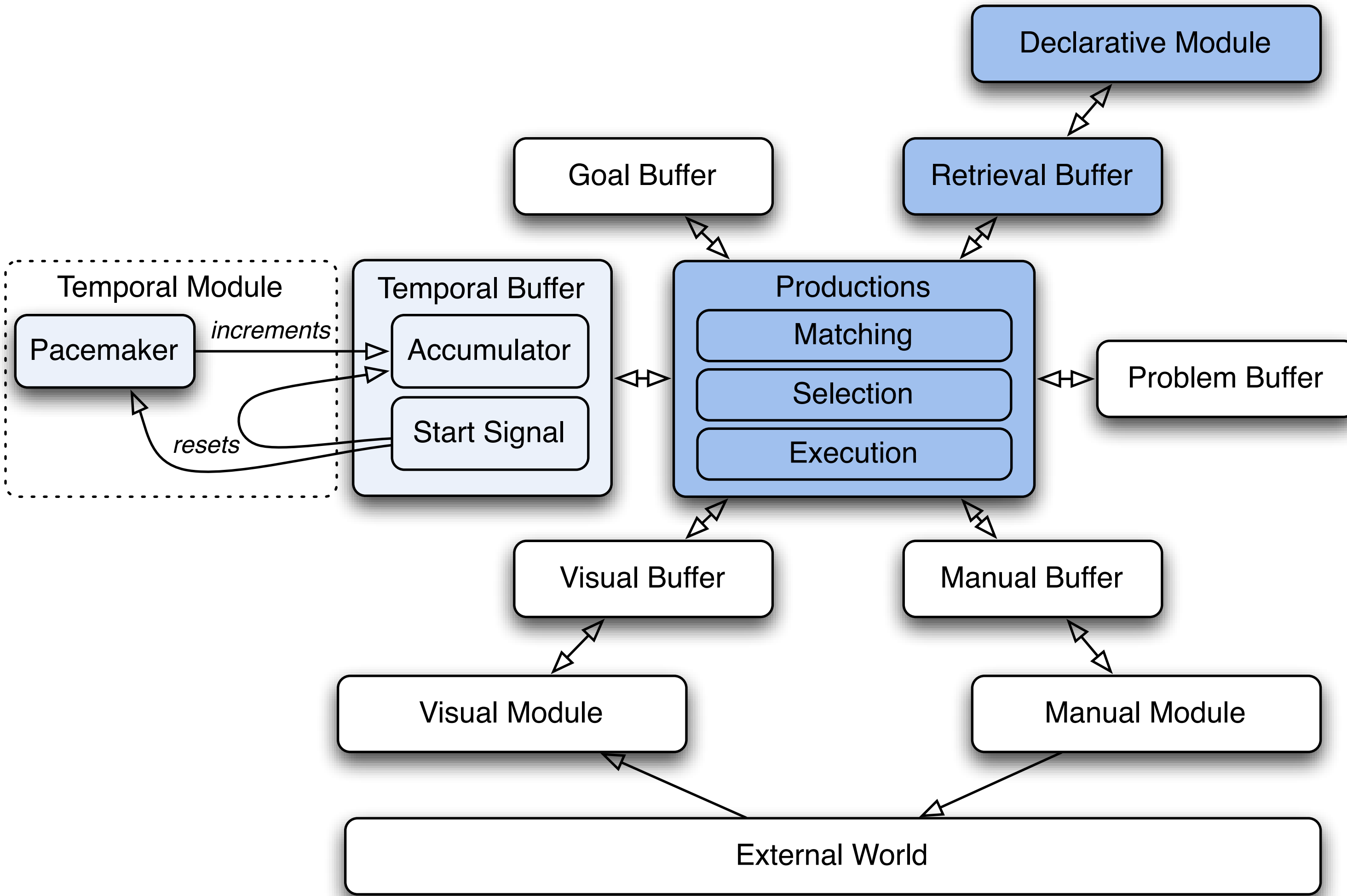


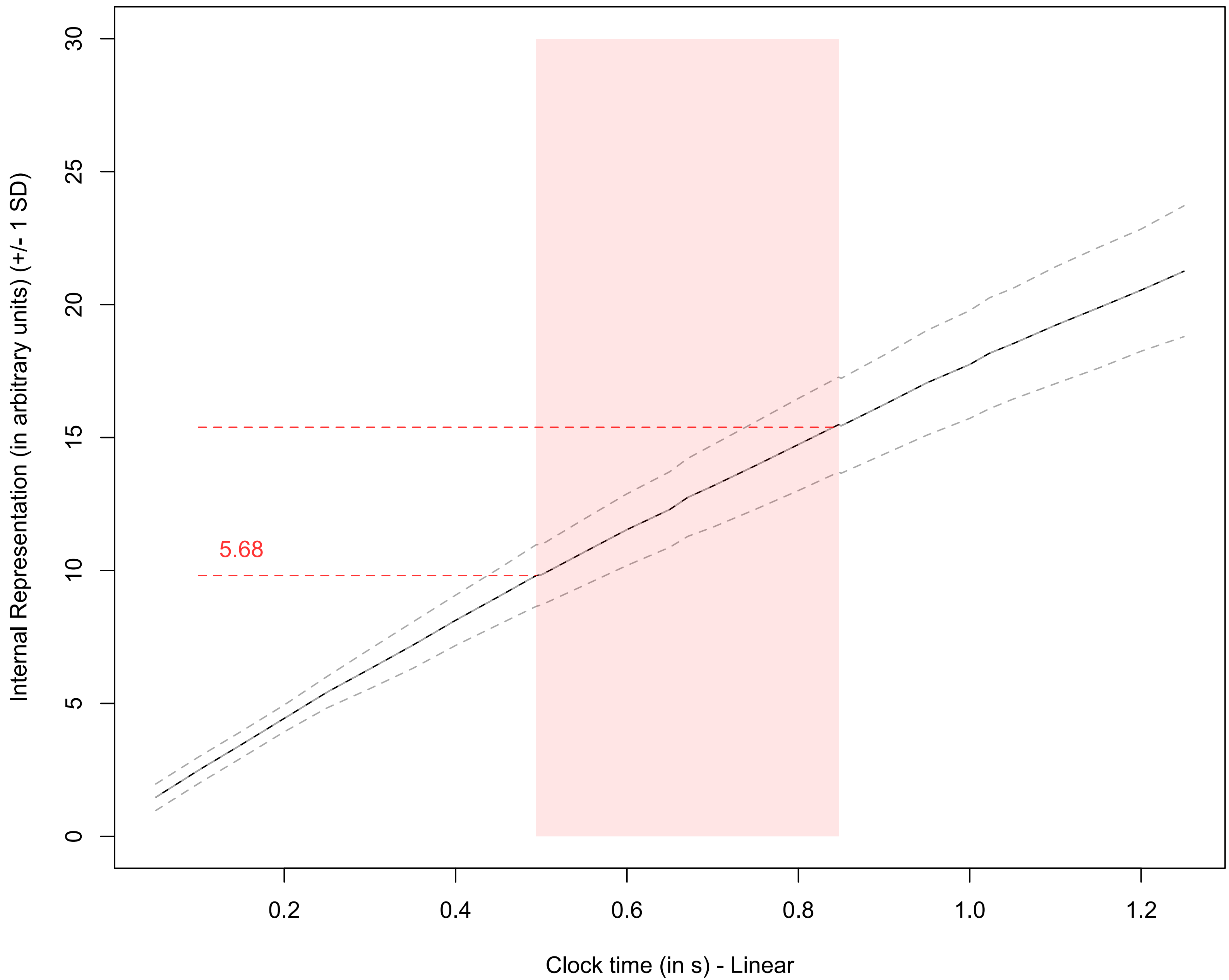
Default "ACT-R" timing model

Blending Model (Dikes & River)

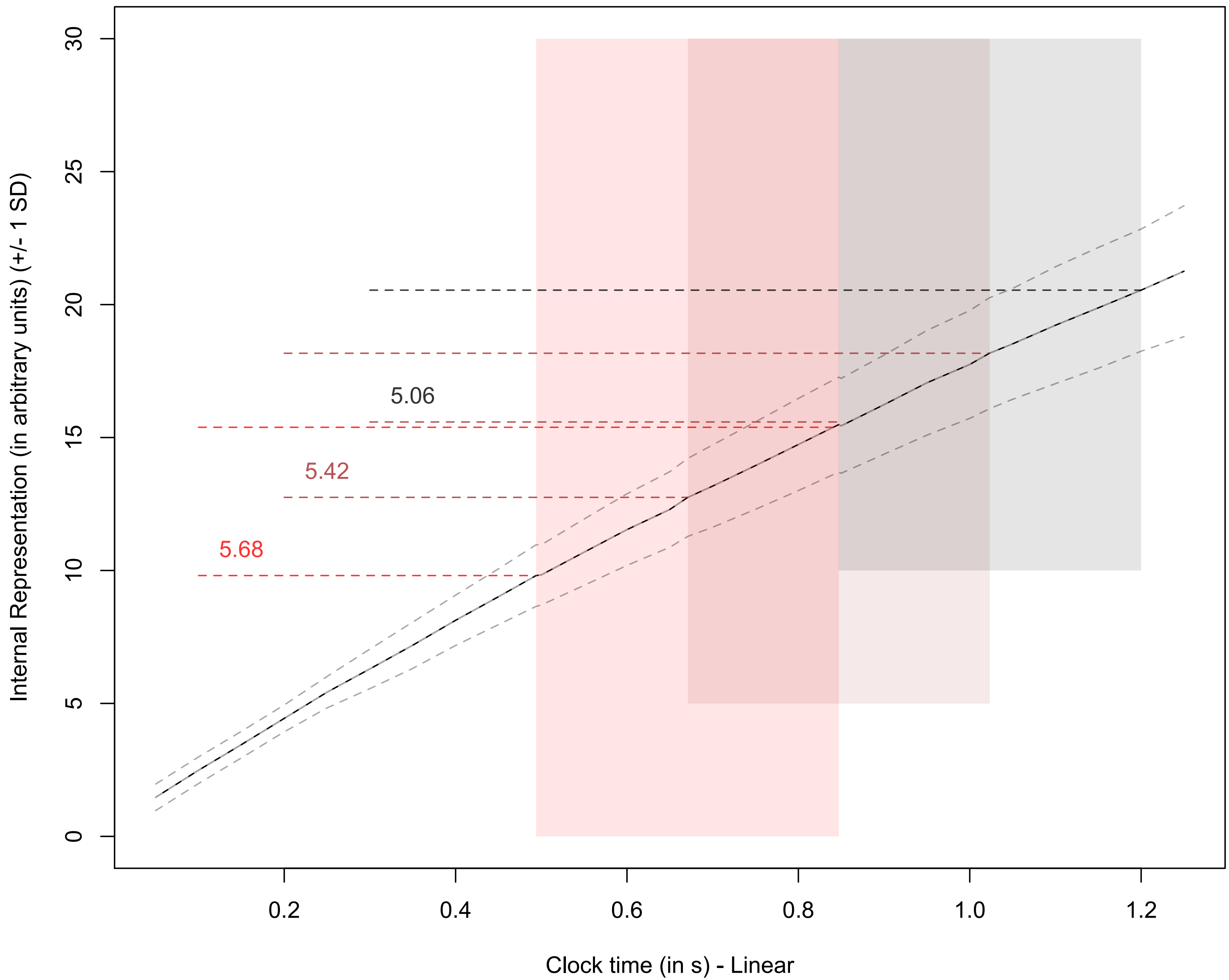
Why not take a much simpler approach?

Taatgen, Van Rijn & Anderson (2007)



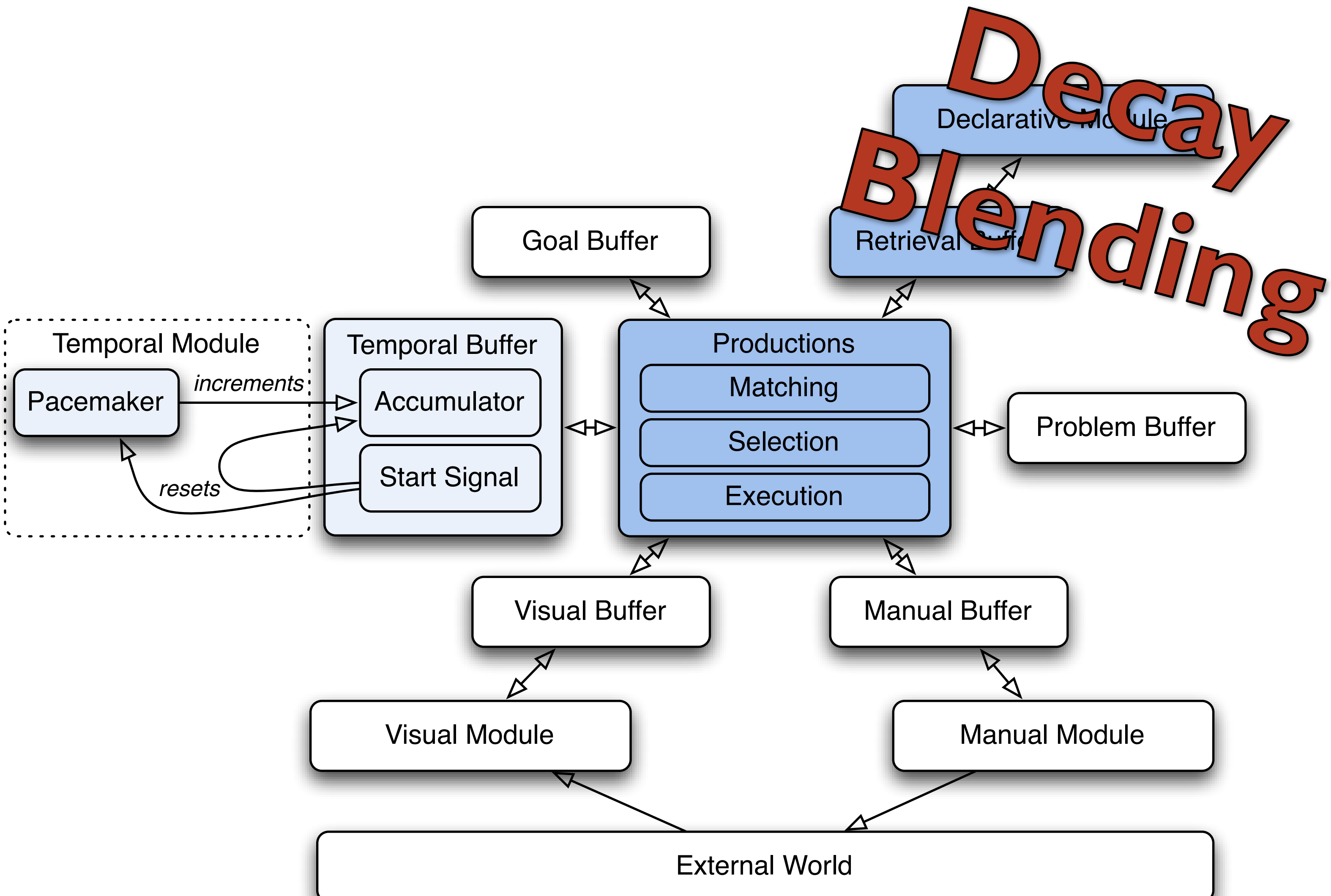


Internal Representation (in arbitrary units) (+/- 1 SD)
0
5
10
15
20
25
30
5.68
0.2
0.4
0.6
0.8
1.0
1.2
Clock time (in s) - Linear



Internal Representation (in arbitrary units) (+/- 1 SD)
Clock time (in s) - Linear
5.06
5.42
5.68
0
5
10
15
20
25
30
0.2
0.4
0.6
0.8
1.0
1.2

Taatgen, Van Rijn & Anderson (2007)



Decay & Blending

$$A_i = \sum_j a_j^{-.5}$$

- Decay:
  - Strength of memory traces decreases with time

A DM account of Contextual Timing

	a_i	V_i	a_i	V_i	a_i	V_i	a_i	V_i
DM	-9	30			-6	34	-3	36

Decay & Blending

- Decay:
  - Strength of memory traces decreases with time

$$A_i = \sum_j a_j^{-.5}$$

$$P_i = \frac{e^{A_i/s}}{\sum_j e^{A_j/s}}$$

A DM account of Contextual Timing

	a_i	V_i	a_i	V_i	a_i	V_i	a_i	V_i
DM	-9	30			-6	34	-3	36
P_i	0.25				0.31		0.44	

Decay & Blending

- Decay:
  - Strength of memory traces decreases with time

$$A_i = \sum_j a_j^{-.5}$$

$$P_i = \frac{e^{A_i/s}}{\sum_j e^{A_j/s}}$$

- Blending:
  - Information retrieved depends on the activation of all traces available in memory; weighted average.

$$t_e = \sum_i P_i V_i$$

A DM account of Contextual Timing

	a_i	V_i	a_i	V_i	a_i	V_i	a_i	V_i
DM	-9	30			-6	34	-3	36
P_i	0.25				0.31		0.44	
$P_i \times V_i$	7.5				10.5		15.8	

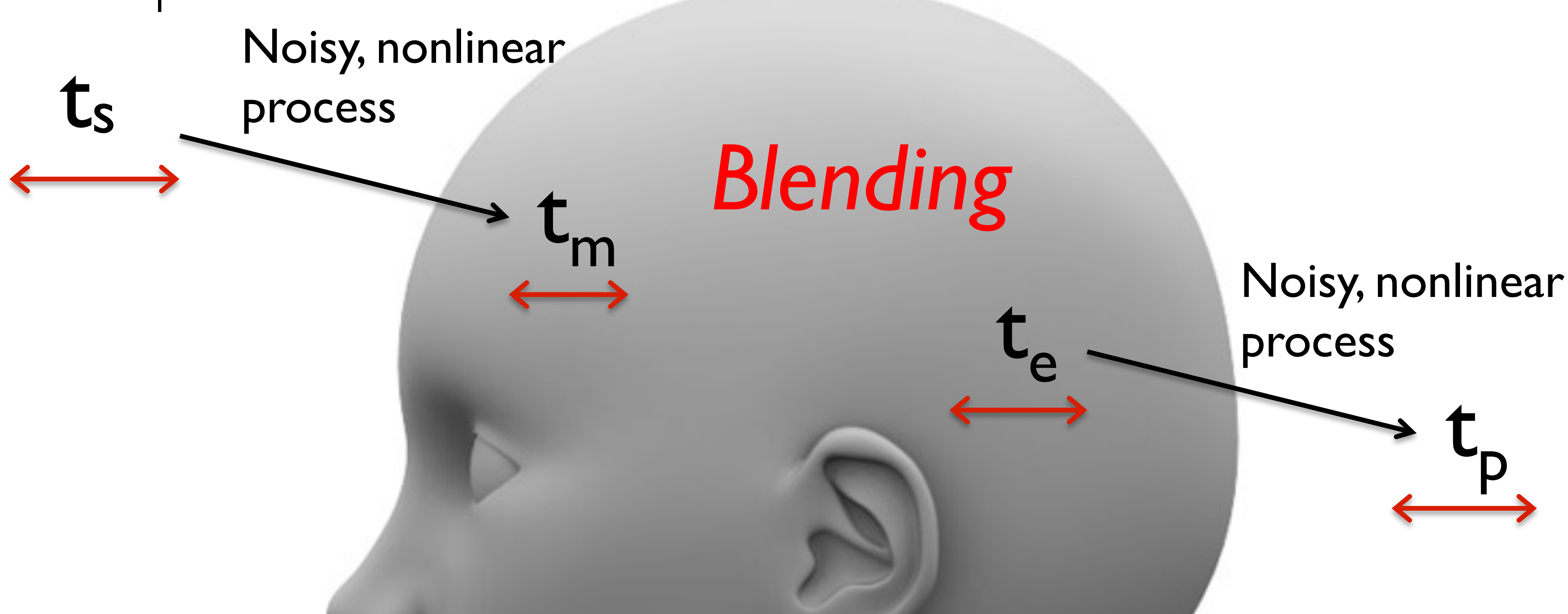
A DM account of Contextual Timing

	a_i	V_i	a_i	V_i	a_i	V_i	a_i	V_i
DM	-9	30			-6	34	-3	36
P_i	0.25				0.31		0.44	
$P_i \times V_i$	7.5				10.5		15.8	

$t_e = 33.8$

ACT-R model of JaS

- Context:
  - 500 practice trials per condition stored in DM
  - Trained with similar step function as in JaS
- Experiment:



A DM account of J&S

	"Prior" a_i	V_i	a_i	V_i	"Prior" a_i	V_i	"Prior" a_i	V_i
DM	-9	30			-6	34	-3	36

A DM account of J&S

	"Prior"		"Likelihood"		"Prior"		"Prior"	
	a_i	V_i	a_i	V_i	a_i	V_i	a_i	V_i
DM	-9	30	-1	31	-6	34	-3	36

A DM account of J&S

	"Prior"		*"Likelihood"*		*"Prior"*		*"Prior"*	
	a_i	V_i	a_i	V_i	a_i	V_i	a_i	V_i
DM	-9	30	-1	31	-6	34	-3	36
P_i	0.14				0.18		0.25	

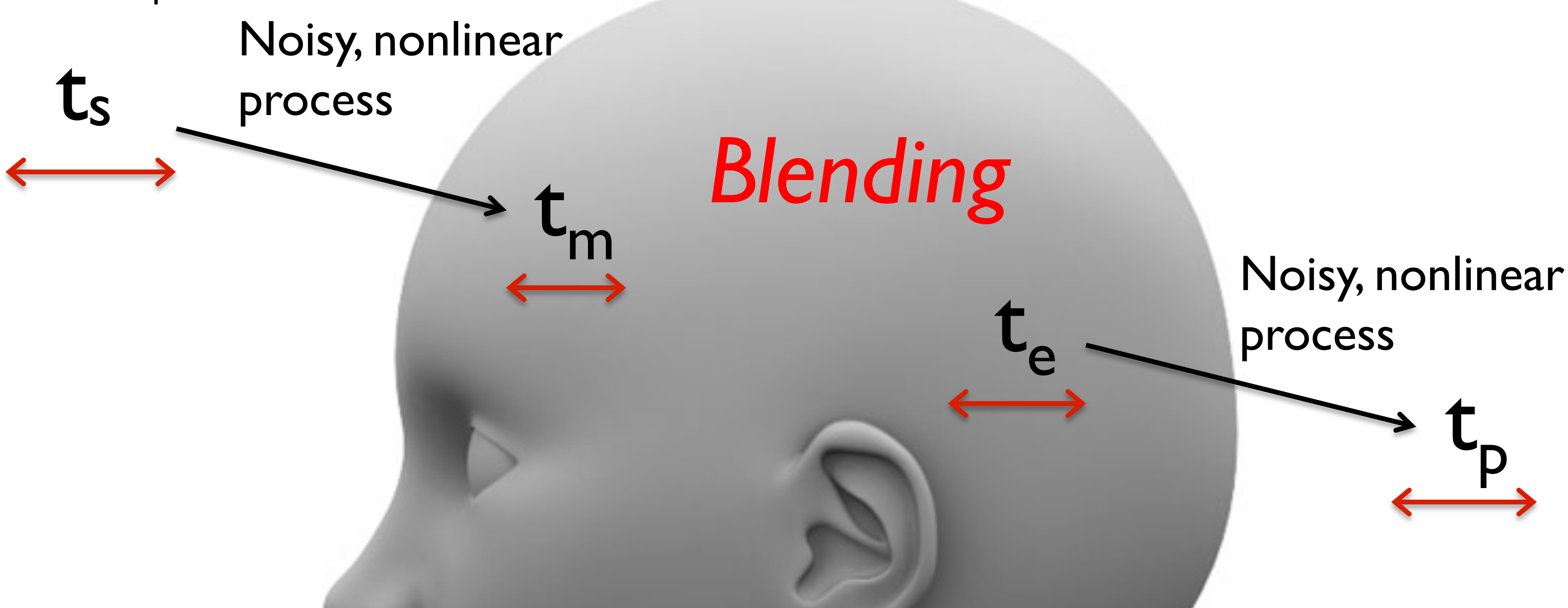
A DM account of J&S

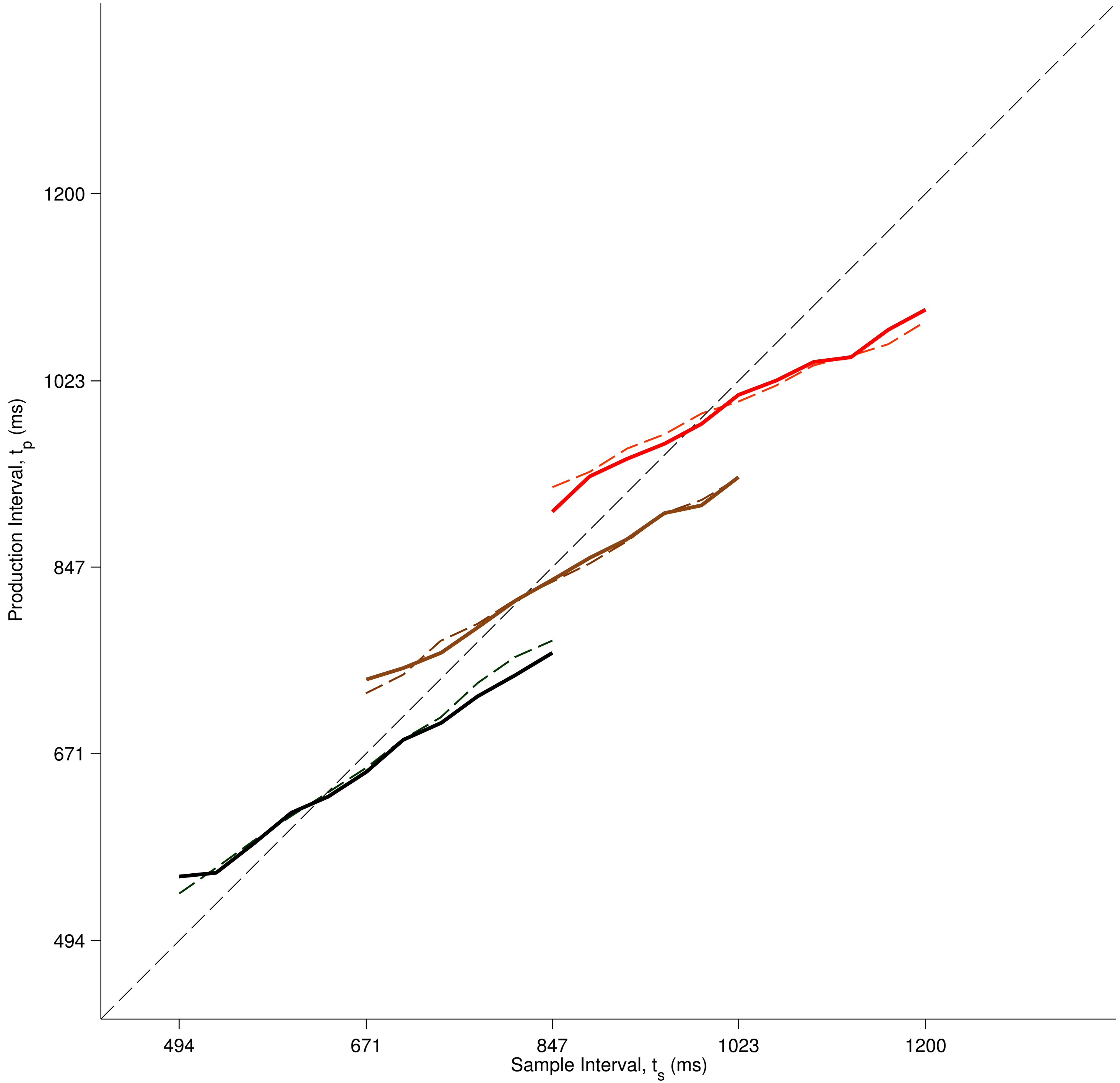
	"Prior"		*"Likelihood"*		*"Prior"*		*"Prior"*	
	a_i	V_i	a_i	V_i	a_i	V_i	a_i	V_i
DM	-9	30	-1	31	-6	34	-3	36
P_i	0.14		0.43		0.18		0.25	
$P_i \times V_i$	4.2		13.3		6.1		9	

t_e=32.6

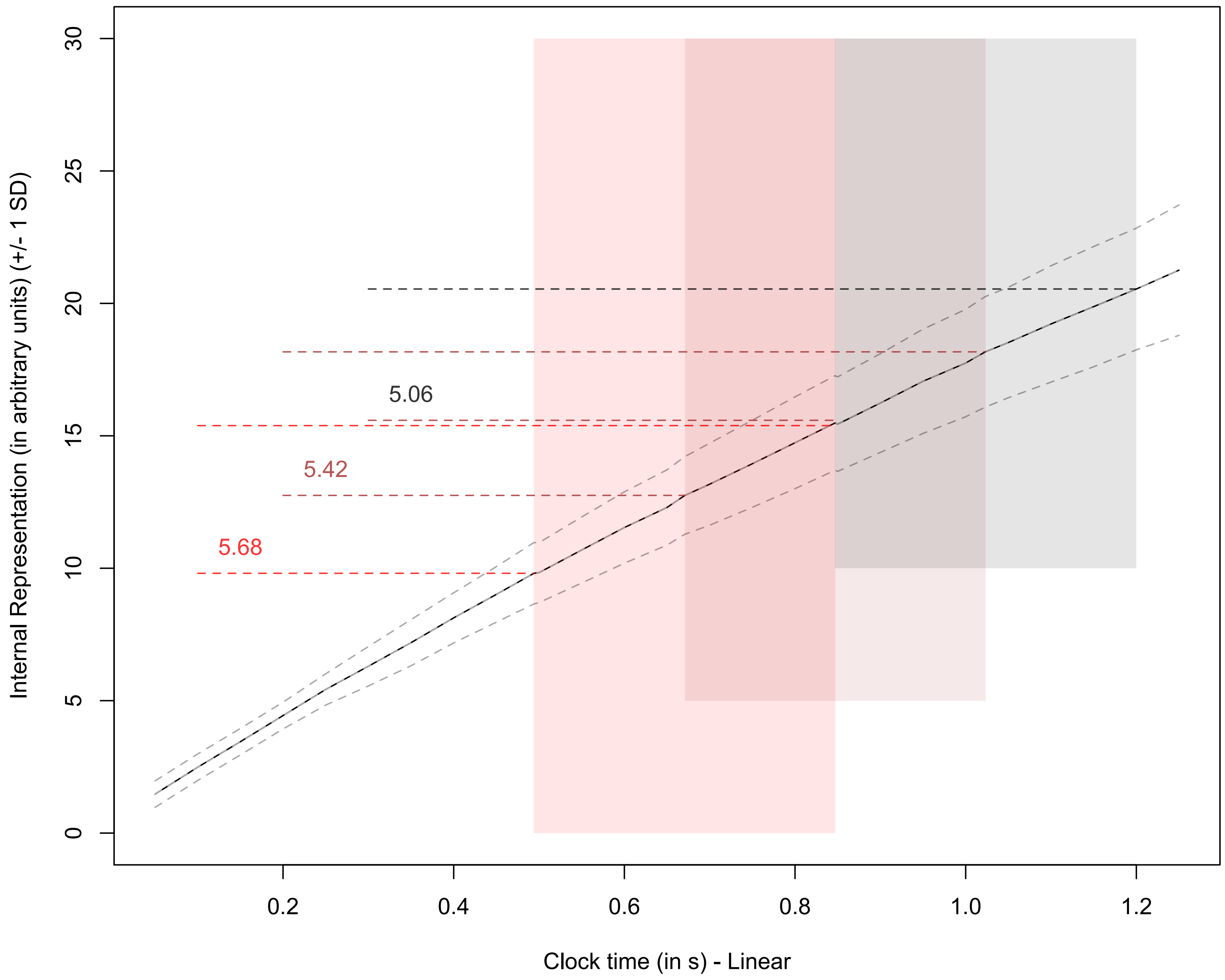
ACT-R model of JaS

- Context:
  - 500 practice trials per condition stored in DM
  - Trained with similar step function as in JaS
- Experiment:

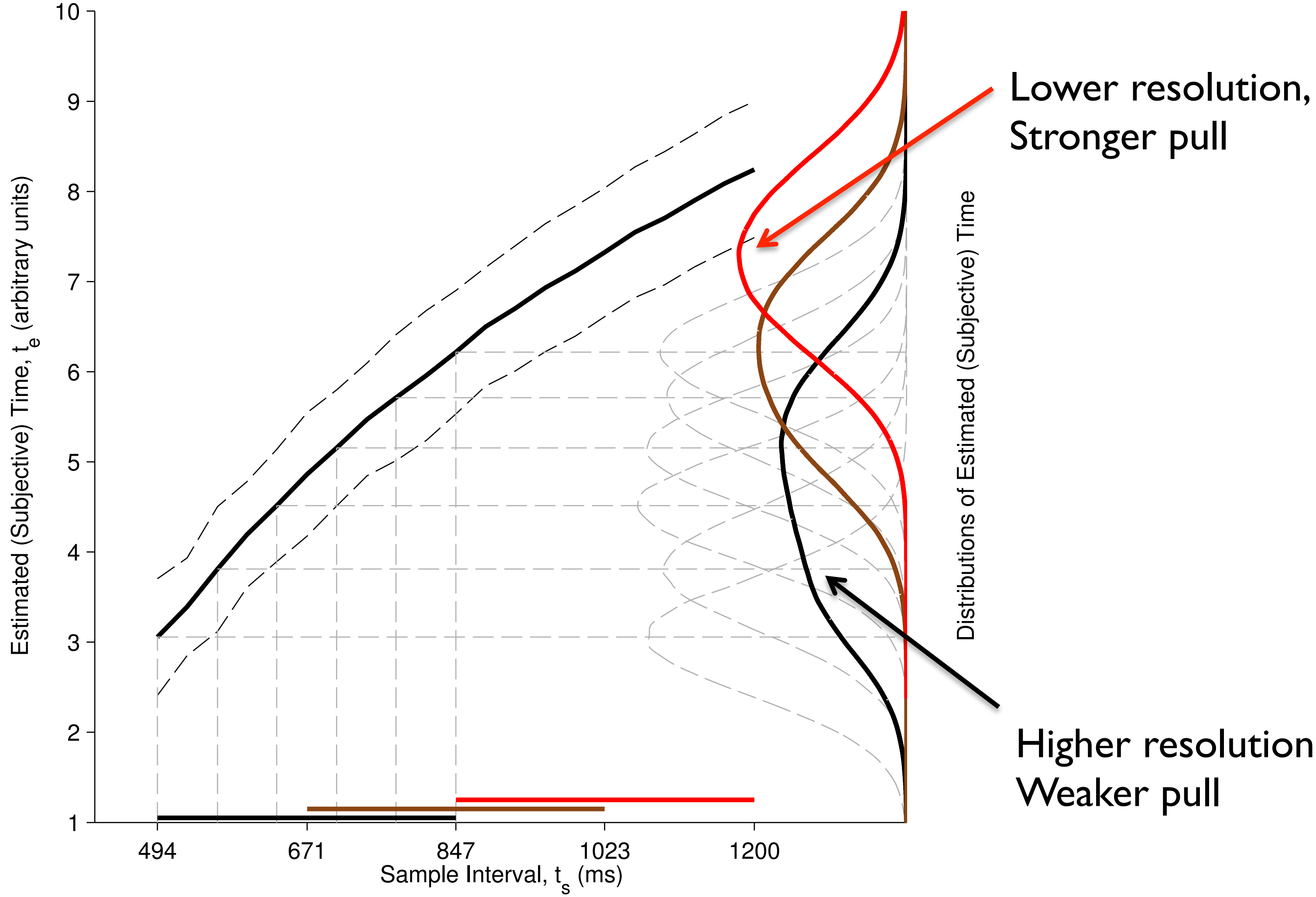




1200
1023
847
671
494
Production Interval, t_p (ms)
494
671
847
1023
1200
Sample Interval, t_s (ms)



Internal Representation (in arbitrary units) (+/- 1 SD)
Clock time (in s) - Linear
0
5
10
15
20
25
30
0.2
0.4
0.6
0.8
1.0
1.2
5.06
5.42
5.68



10
9
8
7
6
5
4
3
2
1
Estimated (Subjective) Time, t_e (arbitrary units)
494
671
847
1023
1200
Sample Interval, t_s (ms)
Distributions of Estimated (Subjective) Time
Lower resolution,
Stronger pull
Higher resolution,
Weaker pull

It's the prior, stupid!

- Likelihood-based explanation:
  - Wider likelihood distributions for longer t_m.
  - Requires the prior to be more constrained than empirically plausible
- Prior-based explanation:
  - Less precise measurements for longer t_s.
  - Non-linear time results in narrower distributions for longer durations, resulting in stronger "pull".

Questions

STT vs Bayesian Models

Assuming the Bayesian inference theory of time perception is right, would we not still need a system for generating pulses to get the probability that a certain amount of time passed? So instead of being the Pacemaker, switch and accumulator, would the likelihood not only be the acuumulator?

In the section "Integrating Bayesian inference with scalar timing theory" the authors venture to integrate two throries on contextual calibration for timing. It seems that the mapping of elements of the Bayesian framework onto elements of Scalar timing theory seems a good fit, but what other reason do the authors have to link these theories? By their own admission the theories have some key differences. In my mind this would require either one solution to be declared the best, or a new theory aspect (such as a decision rule) to be created as the best of both worlds.

The theories are different enough, it seems, to be combined. But where are the clues that they truly cooperate well?

Despite of the 3 main differences between the Scalar Timing Theory framework and the Bayesian perspective, from this paper the overall idea I got is that these are only the two sides of the same coin. So in order to provide new insights on how timing perception works I don't see how promising this approach is. This is also proven by the fact that the authors their selfs at the end of the paper highlight that even relatively simple discrimination and generalization tasks are still unclear and difficult to explain. So what I'm asking is if although the Bayesian framework is able to provide some quantitatively explanations that the Scalar Timing Theory framework isn't able to do is this the correct way to approach the problem? To me it seems that it's just an integration of an already existing framework that still doesn't provide that much more explanations then the initial one. Maybe it's the initial STT framework itself to be wrong, and researchers should rethink about it and then try to apply the Bayesian approach on that.

"Integrating Bayesian inference with scalar timing theory could also be beneficial in explaining other forms of contextual calibration, including the influences of non-temporal factors (e.g., background intensity, speed/sequence structure)."

How could Bayesian inference explain non-temporal factors? How is this related to interval timing/scalar timing theory?

Are these types of models rich enough?

Shi, Church and Meck (2013) show the impressively straight-forward connection between the classic information-processing model and bayesian reasoning of interval timing. [..] The model is able to deliver explanations for phenomena - tested in extremely artificial and simplified environments – through reasonable and logic application. However, in reality interval timing does not follow a strict learning-judgement pattern, but seems to be a strongly intertwined - mostly unconscious - process in an environment of stimuli with different intensity, complexity, familiarity or emotional value, who all possibly have an influence on our subjective time perception (Matthews, & Meck, 2014).

(I) How truly probable is it then, that such a simple "centralised system model" - which already is dealing with quite complex probability judgements in artificially easy tasks - is responsible for the multitude of different timing events we perceive in a rich and immensely complex environment?

Shi, Church and Meck (2013) show the impressively straight-forward connection between the classic information-processing model and bayesian reasoning of interval timing. [..]

(II) Even if components to better integrate stimulus properties will be considered. Isn't the oversimplification of these processes in such a model structure as proposed by Shi, Church and Meck (2013) in its current form, not just a possibly misguided attempt to achieve explanations at an explanatory level that may just not be suitable to describe this highly versatile process (given that explanations of timing in my opinion should supposedly inherit a high external validity due to its ubiquity in everyday life)?

Audio vs Visual

- Why are drummers also better in the visual domain?
- Could it be that the noise we now model as a 'fixed variance' across subjects is the main factor in this equation? Or at least play a bigger role than we now give it? That this noise becomes less variable as we train calibrations to become less dependent on context, that the expert drummer managed to reduce this noise function to have very little influence? This would mean less overlap in estimations, resulting in less 'gravity' between estimations, meaning a more clear distinction between them.

Bonus Question

- Bonus question: How did they do these super-fancy graphs? Awesome!