Modelling human performance on timing tasks with the SET system

John Wearden
Keele University
Exploring the SET system

- The SET system consists of a pacemaker-accumulator internal clock
- Working and reference memory stores
- Decision mechanisms
Aims of the talk

• I want to give some examples of the application of ideas from SET to human timing
• The applications may or may not be consistent with the application of SET to animal timing: animal studies, and related theory, remain a source of inspiration, but shouldn’t impose restrictions on theorizing about human timing, although they can inform it
In particular

- I want to show how the memory and, particularly, decision process elements of SET can be used to model behaviour on a range of tasks
- But there’s another issue....
Is SET a benign influence for good, or evil empire?
Arbitrariness and testability

- The SET system is complicated, and can be made even more complicated by adding things to it (“biases”, random responding, “exotic” decision rules, and so on)
- Using the SET “erector kit” (with additional bits) fits to more or less any data set can (probably) be obtained
- How can we tell that we’ve done isn’t an arbitrary construction that doesn’t really represent reality?
Answer (?)

• We can see whether the components of the models make “psychological sense”
• We can look for orderliness in parameter values with changes in conditions, and with changes in participant groups (e.g. developmental effects)
• A focus of interest will therefore be on the “behaviour of the model” as much as the “model of behaviour”
• In some cases, failures of modelling may reveal weaknesses in our conception of the process modelled
Modelling with SET

- One of the most important features of SET is that it allows us to develop quantitative models of performance which (a) usually fit data well and (b) have parameter values that are psychologically meaningful.
- A simple example is modelling of the temporal generalization task.
Temporal generalization

• Here, people receive examples of a “standard” duration (e.g. 4 s or 400 ms long), then have to decide whether other “comparison” durations are the same as the standard. Accurate feedback is usually given, or the standard is “refreshed” at the start of each experimental block
• The data shown are the average of 83 participants from 2 conditions, one with a 400 ms standard duration the other with a 4 s standard (with counting prevented by a secondary task)

• YES responses are judgements that the comparison duration was the same as the standard
MCG model

• An SET-based model with theoretically meaningful parameters (reference memory/timing variance and decision threshold value) fits the data well

• The MCG model (modified Church and Gibbon model) proposes that people respond YES when

• $\text{abs}(s^* - t)/t < b^*$
MCG model continued

• \( \frac{\text{abs}(s^* - t)}{t} < b^* \)

• Here, \( s^* \) is a sample from the reference memory of the standard. It has an accurate memory of the standard. It has an accurate mean and some coefficient of variation, \( c \).

• \( t \) is the comparison duration, assumed to be accurately timed

• \( b^* \) is a variable threshold
• The model predicts asymmetrical generalization gradients
• It has two parameters, \( c \), memory or timing variability, and \( b \), threshold
• The MCG model specifies the rule for the comparison and the standard to be “close enough” for the participant to judge them as the same
Example

• An example of the use of the MCG model comes from Wearden, Wearden, and Rabbitt (1997) in a study of timing and ageing.

• The MCG model allows us to decide whether any age-related changes are linked to timing/memory variance changes with age, or threshold (decision) changes, or both.
Wearden, Wearden, and Rabbitt, 1997

- Studied subjects from 60-79 years
- Usual confound between age and IQ was avoided by sampling
- Data could be analysed in terms of age (with IQ constant between age-groups) or IQ (with age constant between IQ groups)
Temporal generalization: Age

Figure 1. Mean proportion of yes responses (i.e., identifications of a presented duration as the 400-ms standard) as a function of stimulus duration for the 60- to 69-year-old (A) and 70- to 79-year-old (B) groups on the temporal generalization task (Experiment 1). The circles represent the obtained data, and the lines (e.g., 60–69 mod) connect points derived from the computer model.
Temporal generalization: IQ
Modelling

• Used MCG model
• Main parameters: reference memory or timing variability (c: bigger is worse); threshold for comparison (b)
Results of modelling

Table 1
Parameter Values From Fits of the Modified Church and Gibbon Model to Data From Experiment I

<table>
<thead>
<tr>
<th>Group</th>
<th>Parameter</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>c</td>
<td>b</td>
<td>x</td>
</tr>
<tr>
<td>60–69 years</td>
<td></td>
<td>.26</td>
<td>.28</td>
<td>.13</td>
</tr>
<tr>
<td>70–79 years</td>
<td></td>
<td>.37</td>
<td>.30</td>
<td>.15</td>
</tr>
<tr>
<td>Low IQ</td>
<td></td>
<td>.50</td>
<td>.33</td>
<td>.15</td>
</tr>
<tr>
<td>Mid-IQ</td>
<td></td>
<td>.30</td>
<td>.31</td>
<td>.15</td>
</tr>
<tr>
<td>High IQ</td>
<td></td>
<td>.24</td>
<td>.26</td>
<td>.12</td>
</tr>
<tr>
<td>Students (50%)</td>
<td></td>
<td>.16</td>
<td>.25</td>
<td>.14</td>
</tr>
<tr>
<td>Students (10%)</td>
<td></td>
<td>.20</td>
<td>.25</td>
<td>.10</td>
</tr>
</tbody>
</table>

Note. c is coefficient of variation of memory of the standard 400-ms duration, b is the threshold mean, x is the threshold standard deviation, and MAD is the mean absolute deviation in the proportion of yes responses between the obtained data and those obtained from the model. The two student groups were participants in Wearden’s (1992, Experiment 2) research.
Conclusions

• MCG model fits well in all cases
• Main effect of age is on temporal memory/timing variability (larger when older)
• Main effect of IQ is on temporal memory/timing variability (larger when IQ lower)
• IQ effect bigger than age effect
Behaviour of the model

• Droit-Volet, Clement, & Wearden (2001) used children of 3, 5, and 8 years
• Standards were 4 and 8 seconds (children of this age don’t count spontaneously)
• When we put all the data together we get...
Effect of age

![Graph showing the effect of age on memory variation. The x-axis represents subject age (years) ranging from 0 to 80, and the y-axis represents the coefficient of variation of memory ranging from 0 to 0.6. The graph shows a decrease in memory variation with age, peaking around the 20-year mark.](image-url)
What this shows

• These examples show how the information-processing version of SET can be used not only to model data from humans accurately, but also to offer theoretical insight into underlying processes.

• Changes with age in the memory/timing variance are orderly and “intuitive”.

• But what about decision processes?
Ferrara, Lejeune, and Wearden, 1997

- People performed temporal generalization with a 600 ms standard.
- In different groups, the non-standard comparison durations were widely-spaced (150 ms), or closely-spaced (75 ms) around the standard.
Results
Things to note

• Stimuli common to the two sets (e.g. 600 and 750 ms) are confused less frequently in the “hard” group than the “easy” one
Interpretation

• Modelling showed that the Ferrara et al. effect (which is very commonly found, with different tasks, and even with rats) can be explained by a change in decision processes

• That is, people become more conservative about saying YES (smaller \( b \)) in the temporal generalization task when the task is more difficult
Replication: Paul et al.

- This was identical to Ferrara et al., 1997, except that the stimuli were visual.
- The study involved ERPs, so each participant received hundreds of trials, far more than in Ferrara et al.
Data

![Graph showing the mean proportion of YES responses as a function of stimulus duration. The graph compares Easy and Difficult conditions, with separate lines for Easy simulation and Difficult simulation.](image-url)
Parameter values

- c was 0.36 in both conditions
- b was .36 for the easy spacing and .29 for the difficult spacing
- Given that the stimuli were the same on average, you’d expect c to be the same (and it was), but the decision process is more conservative in the difficult case, and this affected only b in the model
“Episodic” temporal generalization

- Another task which illustrates “difficulty” effects, and which also shows how new data can be modelled with SET is “episodic” temporal generalization
- This is essentially a temporal generalization experiment without any reference memory
Procedure

• On each trial, people receive two stimuli (e.g. tones), and they have to judge whether or not they have the same duration

• One of the stimuli (a “standard”) is drawn randomly from a distribution (e.g. 400-600 ms), and the other (a “comparison”) is some multiple of it (e.g. from .25 to 1.75)

• The order of the stimuli is randomized
Things to note

• The stimuli are never repeated, except by chance, so (presumably) no “reference memory” is used

• Data can be plotted in terms of the “comparison”/”standard” ratio

• The next slide shows some data from conditions where the multiples were varied (to make an “easy” and “difficult” set of pairs)
Episodic temporal generalization

![Graphs showing mean proportion of YES responses against comparison/sample ratio for easy and difficult conditions.](image_url)
Superimposition graph
Things to note

• The gradients were slightly asymmetrical

• The gradients from conditions with different levels of “difficulty” don’t superimpose (cf. Ferrara et al., 1997)

• If there’s no reference memory, how can the data be modelled?

• The simple MCG model can’t be used, as this assumes a difference in principle between the standard and comparison durations
Model

- Respond YES when
- \[ \frac{\text{abs}(s_1^* - s_2^*)}{\text{mean}(s_1^*, s_2^*)} < b^* \]
- \( s_1^* \) and \( s_2^* \) are the two stimuli (represented with scalar variance) and \( b^* \) is a (variable) threshold
- Does the model work?
YES!

Comparison/sample ratio

Mean proportion of YES responses

- .25 SHORT
- .25 LONG
- SHORT MODEL
- LONG MODEL

Comparison/sample ratio

Mean proportion of YES responses

- .125 SHORT
- .125 LONG
- SHORT MODEL
- LONG MODEL
Things to note

• The gradients are asymmetrical even though there’s no “reference memory” component
• The modelling illustrates how an extension of the MCG model can be used to model data from a new task
• The forms of the MCG and the EGM (episodic generalization model) are both “thresholded normalized differences”
• That is, the numerator is a difference between relevant quantities, this is divided by something to scale it, and the result is compared with a threshold
Developmental studies with children

- Variants of the MCG model have been used to simulate data from children
Droit-Volet, Clement, and Wearden, 2001

• Temporal generalization in children of 3, 5, and 8 years
• Standards were 4 and 8 seconds (children of this age don’t count spontaneously)
3 year-olds

![Graph showing the mean proportion of YES responses for 3 year-olds with stimulus duration (seconds) on the x-axis and mean proportion of YES responses on the y-axis. The graph includes two lines: one for 4 seconds and another for 8 seconds. The line for 4 seconds starts at a lower proportion, rises to a peak around 6 seconds, and then decreases. The line for 8 seconds starts higher, peaks at about 8 seconds, and then decreases.]
5 year-olds
8 year-olds
Empirical results

- Gradients were flatter in younger children than older ones
- Gradients were symmetrical in 3 and 5-year-olds but asymmetrical in 8 year-olds
Modelling

• The standard human temporal generalization model was used, with the addition of (a) a “random responding” parameter and (b) “distortion”, k: <1 duration remembered as too short; >1 remembered as too long

• So, the rule for responding YES becomes, respond YES if abs(ks* - )t/t < b*
Modelling parameters

Parameter values from modified Church and Gibbon model with random responding and standard memory distortion

<table>
<thead>
<tr>
<th>Age</th>
<th>Standard (s)</th>
<th>c</th>
<th>b</th>
<th>p</th>
<th>k</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4</td>
<td>.59</td>
<td>.19</td>
<td>.52</td>
<td>.83</td>
<td>.03</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>.53</td>
<td>.22</td>
<td>.39</td>
<td>.88</td>
<td>.02</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>.55</td>
<td>.19</td>
<td>.36</td>
<td>.90</td>
<td>.03</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>.57</td>
<td>.27</td>
<td>.22</td>
<td>.83</td>
<td>.03</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>.35</td>
<td>.29</td>
<td>.12</td>
<td>.90</td>
<td>.02</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>.36</td>
<td>.26</td>
<td>.12</td>
<td>1.0</td>
<td>.02</td>
</tr>
</tbody>
</table>

Note: The model parameters are: c = coefficient of variation of the memory representation of the standard; b = mean threshold value; p = probability of random responding on each trial; k standard memory distortion value. The standard deviation of the threshold was always b/2. Also shown is mean absolute deviation (MAD).
“k”: nice....but naughty.....

• The data can be modelled using the idea that the youngest children are remembering the standard as shorter than it really is

• The change is orderly with age

• But why should young children do this at all?
Ogden et al. 2008: JEP:HPP

• This was a very complex experiment which essentially demonstrated (or seemed to demonstrate) interference between one memorized temporal generalization standard and another one.

• Temporal generalization gradients were sometimes very distorted, so peaked at comparisons remote from the standard.
Sample data

![Graph showing the mean proportion of YES responses against the comparison/standard ratio for different conditions: Shorter (filled circles), Same (open circles), and Longer (filled triangles).]
Modelling

• The distorted (and normal) temporal generalization gradients could be modelled by changing $k$: other parameters ($c$ and $b$) didn’t change systematically.
Modelling

![Graph showing comparison of different models with varying comparison/standard ratios and mean proportion of YES responses.](image)

- Shorter
- Same
- Longer

- Shorter model
- Same model
- Longer model
In this case....

• ....the use of what’s supposed to be a memory parameter, like k, doesn’t seem too bad, as the experiment involves interference which should be a memory manipulation, but using it to explain between-group differences is more problematical, and raises the question of why group A and group B should differ in this respect

• I’ll provide another example later
Temporal bisection

• Another commonly-used task is **bisection**
• Potentially there’s a lot to say here, but I just want to use bisection to illustrate a case where the “behaviour “ of models was found wanting
Basic method: classification

• Derives from animal experiment by Church and DeLuty (1977)

• People receive examples of Short ($S$) and Long ($L$) standards, e.g. 200 and 800 ms tones

• They then receive other tones (200-800 ms in 100-ms steps) and classify each in terms of similarity to $S$ and $L$
Resulting data

• The plot of proportion of “Long” responses (i.e. judgements that a duration is more similar to $L$ than to $S$) against duration forms a \textit{psychophysical function}
Psychophysical functions

![Graph showing psychophysical functions with stimulus duration on the x-axis and mean proportion of LONG responses on the y-axis. The graph includes two lines with different markers to represent different conditions or subjects.](image-url)
Intuitive modelling?

• Suppose that we have S and L, and some test value, t
• Calculate the difference between t and both S and L
• \((L - t)\) and \((t - S)\)
• these are equal when \(2t = (L + S)\), so
• \(t = (L + S)/2\), the arithmetic mean
• Wearden (1991) complicated the rule a bit by supposing that if the two differences were similar the model responded “Long”
Basic idea

• The basic idea behind many models of bisection is that the decision rule is of the form $R = f(S, L, t)$: what exact decision rule is proposed varies between models.

• That is, the response (Short or Long) depends on the values of $S$ and $L$ (which are stored in reference memory) and the value of $t$, in working memory.

• But....then there was ...
Wearden and Ferrara, 1995

• We introduced a new method (“Partition bisection”)

• People received a set of tones in a random order (e.g. 200, 300,…..800 ms), and just had to classify them as “Short” or “Long”

• After a number of repetitions, the psychophysical function looked “normal” in all respects
• So maybe $S$ and $L$ have no special status
• W & F also looked at stimulus spacing effects between a constant $S$ and $L$ (e.g. linear versus logarithmic), and found that log spacing shifted the psychophysical function to the left
Linear versus log spacing

against stimulus duration, from subjects in Experiment 1 using the 200/800-msec sets, the right panel data from the 100/
Unequal spacing

FIG. 5. Proportion of Long responses plotted against stimulus duration from subjects in Experiment 2. The left panel shows data from the 200/800-msec duration sets, the right panel data from the 100/900-msec sets. Open circles show data from the appropriate SIH group, filled circles data from the appropriate LNG group.
Conclusions

• S and L may have no special status
• spacing between S and L makes a difference, so.....
• A model just based on S, L, and t can’t work
• So the “behaviour of the model” is unsatisfactory, even though it may fit the original data set well
Proposed model

• Bisection is based on comparisons between \( t \) and the mean of all the durations presented, \( M \)

• \( S \) and \( L \) have no special status, they just contribute to \( M \)
Rules

- Respond “Long” when $(t - M)/t > b$
- “Short” when $(t - M)/t < -b$
- at random when $-b < (t - M)/t < b$
Things to note

• The rule is much more similar to the temporal generalization rule

• There’s a threshold as well (so the overall model is of the “thresholded normalized difference” type)

• Does it fit?
It fits the LIN and LOG results
...and the unequal spacing results

FIG. 7. Proportion of Long responses obtained experimentally in Experiment 2, and predicted by the model. The left panel shows 200/800 conditions (open circles SHT; filled circles LNG), with the solid line the model's output for LNG conditions, the dotted line its output for SHT. The right panel shows analogous measures when the set ranges were 100/900 msec.
So....

- Wearden and Ferrara’s model fitted the data well, as was also more theoretically consistent with temporal generalization decision processes than other work but....
- ....it’s largely been ignored
k again: Allman, DeLeon, and Wearden, 2011

• This was an experiment using bisection, with $S$ and $L$ values of 1 and 4 s, or 2 and 8 s

• Participants were typically developing or had autism spectrum disorder
Results
Comments

• ASD participants had bisection functions shifted to the left relative to typically-developing participants

• This could be modelled by assuming differences in \( k \), the memory distortion of S and L: i.e. \( k \) was systematically lower in the ASD group

• Ok, that fitted, but why should this be true?
Story so far

- You’ve seen how models inspired by SET, which fit data well, can be developed
- The models can be tested by comparing results of experimental manipulations with what would be expected (intuitively?) of the models’ parameters
- Most applications seem “principled”, but the use of “memory distortion” for between-group comparisons raises problems (at least in my view)
I want to talk about models of some classical timing phenomena

• Reproduction
• Verbal estimation
• Chronometric counting
Reproduction

• In reproduction, people receive some target duration (usually in the form of a stimulus), $s$, and have to make some response with the same duration, often by stopping another stimulus when the “time is up”

• Reproduction often (but not always) obeys Vierordt’s Law: short durations are reproduced as longer than they are, whereas long durations are reproduced as shorter than they are
Lejeune and Wearden (2009)

• We presented examples of data from Vierordt’s 1868 monograph
• We also reviewed the more recent literature on Vierordt’s Law
Typical reproduction data
Why do these effects occur?

• One model, for short interval reproduction at least, is to assume a two-stage sequence

• The person stores $s$, the sample, correctly, and times from zero until the current time, $t$, is “close enough” (i.e. at $bs$, where $b < 1$). Then the response is initiated, and this takes some time $d$ to be emitted,

• The total reproduction, $r = bs + d$

• $d$ was assumed to be on average 250 ms
Reproduction model

![Graph showing reproduction model with different thresholds: 10%, 20%, 30%, and 40%. The graph plots reproduction/sample time against sample time (milliseconds).]
The model....

• ...simulates the rudiments of Vierordt-like effects in reproduction, at least in this short range

• The basic time representation has mean accuracy and scalar variance, although the reproductions are not accurate

• Now we turn to a (much) more complicated case: *Verbal estimation*
Verbal estimation

- By verbal estimation I mean the procedure of assigning verbal labels (in conventional time units like seconds or milliseconds) to events, usually the duration of stimulus presentations.
- In all the cases that I’m going to discuss, this assignment doesn’t depend on counting, as the time intervals are too short.
Verbal estimation

• The technique is particularly useful, as it can provide estimates from a range of short stimulus values (e.g. 10 to 15-fold or more, from 50+ ms to 1000+ ms), without the involvement of counting.

• Having a wide range is useful for testing various theories which predict multiplicative effects (e.g. clock speed).
Use of verbal estimation

- Verbal estimation has been used in
- “speeding up the clock” studies
- “tones and lights” studies
- “filled and unfilled intervals” studies
- In all cases it seems to be sensitive to small changes in subjective time
General problem

• Although verbal estimation is used a lot, we have no model of performance on this task (i.e. no idea how participants are doing it)

• This usually isn’t important when different stimulus types (e.g. auditory and visual) are compared...

• ...but it would be nice to know!
More specific problem

• There’s evidence that data from verbal estimation studies violate the “scalar properties” of time
• The scalar properties are
• **Mean accuracy**: the average “estimate” of some time, \( t \), is on average exactly equal to \( t \)
• **Scalar variance**: the standard deviation of “estimates” is a constant fraction of the mean, implying a constant coefficient of variation \((sd/mean)\) as the timed interval changes
Verbal estimation data
The data set you’ve just seen

- Are the average of 80+ participants estimating tones with durations from 77 to 1183 ms
- The mean versus real time function is linear, but not completely accurate
- The coefficient of variation generally declines with increasing duration value, but is smaller at the shortest value. The turndown at the shortest value isn’t always found, but the decline in cv nearly always is
Modelling verbal estimation

- Casual observation of data from participants suggests that the output values are “quantized”, that is, only certain values are used.
- These mostly end in “00” (100, 500, 1000), but sometimes values ending in “50” (particularly 250 and 750) are used.
Questions

• Does the quantization produce the effects observed in average estimates?
• How can the process be modelled quantitatively?
Attractor model: 1

- The attractor model is based on 3 simple principles.
- Stimulus durations are represented “fuzzily”, so don’t give rise to the same “raw sensation” each trial. Some stimulus value, $s$, is replaced on each trial by $s^*$ which is a sample from a Gaussian distribution with mean, $s$, and some coefficient of variation, $c$. This is essentially the scalar property of time in the “raw sensations”
Attractor model: II

• For some particular “raw sensation”, $s^*$, the output value is determined by a set of “attractors”, each of which has a value which determines output (500, 1000, etc.), and also a weight which determines how strong the attractor is
Attractor model: III

- For some value, $s^*$, all the attractors in the set exert a “force” which depends on the weight of the attractor divided by the “temporal distance” from the attractor value to $s^*$
- The two “strongest” compete for output, with the probability of a each value being determined by its relative weight
- So, if one of the two strongest attractors for $s^*$ has twice the attraction “force” of the other one, the probability of this value being output is 2/3, and so on
Parameters

• The model has a potentially infinite number of parameters, but in practice usually fewer than 20

• One of these is the coefficient of variation of the underlying time representations

• The others are (a) the attractor values and (b) the attractor weights

• Obviously, without the latter, the model cannot be run
Some preliminary results

• The model was embodied in a Visual Basic program

• It estimated stimulus durations of 77, 203, 348, 461, 582, 767, 834, 958, 1065, and 1183 ms

• The underlying time representations were scalar (i.e. slope = 1, constant cv)

• The cv value in the simulations was 0.2
Simulation details

• 10000 trials were run at each stimulus duration
• The attractor set, and weights, are shown on the next slide
• The model used output values of 50, 100, 250, 500, 750, 1000, and 1250 ms, but these were not equally weighted (e.g., 100, 500, and 1000 were the strongest)
Attractor set used
Simulation results

![Graph showing mean simulated estimate vs. stimulus duration (ms)](image)

- Mean verbal estimate (ms): 600, 800, 1000, 1200
- Coefficient of variation: 0.3, 0.4, 0.5

![Graph showing coefficient of variation vs. stimulus duration (ms)](image)
Comments

• Although the underlying representations of time are scalar, the output clearly violates scalar properties, particularly with respect to coefficient of variation

• Coefficients of variation generally decline with increasing duration value, but the shortest stimulus duration has a lower cv than the next longest...just like the data

• The attractor model might be considered to be a very complicated decision process in the SET system
Chronometric counting

• Chronometric counting is a very common behaviour in timing experiments with humans, for intervals longer than just over one second

• Data obtained from studies where counting is permitted almost always violate the scalar property of time, in that timing becomes more sensitive as the interval lengthens, suggesting a Poisson-like process
Killeen and Weiss, 1987

• Provided an extensive mathematical analysis which showed why counting produces non-scalar variance in very general cases, but did not provide a model of how chronometric counting occurs on any particular task
Counting in temporal generalization

• Participants counted on a temporal generalization task, with a 4 s or 8 s standard, with counts recorded by key-presses
• At the end of the comparison stimulus, people judged whether or not it was the standard
• Individual counts were measured, as were (a) start times, and (b) whether the stimulus finished before the last count was made, or after it
Basic model

• The basis of the model was the simple idea that (a) a “pool” of numerical values (which could contain more than one value) represented the standard and (b) if the final count at the end of the stimulus was a number in the pool, a YES response was generated.

• The counts themselves had scalar properties of variance.
Counting model

Count generation → End of trial count

Standard pool → Decision process

Comparison count → Decision process

Decision process

YES → NO
Data and model

![Graph showing the mean proportion of YES responses vs. stimulus duration for data and model.]

- **Data and Model 4 s**
  - Stimulus duration (seconds): 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5
  - Mean proportion of YES responses: 0.0, 0.2, 0.4, 0.6, 0.8, 1.0

- **Data and Model 8 s**
  - Stimulus duration (seconds): 4, 5, 6, 7, 8, 9, 10, 11
  - Mean proportion of YES responses: 0.0, 0.2, 0.4, 0.6, 0.8, 1.0
Comment

• The model correctly produced the non-scalar nature of the YES response distributions, even though the underlying counts had a standard deviation than was proportional to their mean (i.e. were scalar)

• The model showed that (a) mean count rate made little difference but (b) variability in mean count between one trial and another was the most important determinant of variability
Conclusions

• You’ve seen how models of varying complexity which incorporate underlying scalar representations of time can model timing data which do not have scalar properties.

• The verbal estimation and counting models are complex by the usual SET standards, but are potentially testable (e.g. by changing the “permitted” responses in verbal estimation).