III. Mathematical modeling of behavioral and neuroimaging numerical tasks

I. Behavior

I.1. The logarithmic number line model: basic formalism

Analog models of number processing assume that each numerical quantity is represented internally by a distribution of activation on an internal 'number line' (Dehaene, 1992; Shepard, Kilpatrick & Cunningham, 1975; Van Oeffelen & Vos, 1982). This internal representation is inherently noisy and varies from trial to trial. Assuming a specific form for this representation, psychophysical tools can then be used to evaluate the optimal strategy and expected success rate (Green & Swets, 1966).

In agreement with previous theories, we adopt the hypothesis of a logarithmic internal number line with fixed Gaussian variability. Note, however, that the alternative hypothesis of a linear scale with linearly increasing or "scalar" variability gives virtually equivalent results, with the exception of subtle asymmetries in the data.

Mathematically, the numerosity of a set of n dots is represented internally by a Gaussian random variable R with mean Log(n) (natural logarithm of n) and with standard deviation w.

$$p[R \in [r, r + dr)] = \frac{e^{-\frac{(r-\mathrm{Log}[n])^2}{2w^2}}}{\sqrt{2\pi}\, w}\, dr$$

In this equation, w is the *internal Weber fraction* that specifies the degree of precision of the internal quantity representation. How should one interpret concretely a given value of w ? Because w expresses an amount of jitter on a logarithmic scale, it can be transformed by exponentiation into a characteristic discriminability ratio $d = e^w$. The latter is a concretely graspable quantity, the typical ratio between two numbers that corresponds to one standard deviation on the internal number line. When w is much smaller than 1, d and $1+w$ are almost equal. This implies that w itself can be readily interpreted as a characteristic numerical deviation. Most intuitively perhaps, $2\,w$, when expressed as a percentage, gives approximately the 95% interval for internal samples representing a given number.

To render this discussion more concrete, our psychophysical data in human adults give $w = 0.17$. This means that two numbers that differ by about 17% (e.g. 6 and

7, or 100 and 117) are just within one standard deviation of the internal variability. It also means that the 95% confidence interval of the internal representation of a given number falls within 34% of that number -- which means that a set of 100 dots is unlikely to be confounded with a set of 134 dots.

Note that w is a hidden, abstract variable that characterizes an internal representation. Thus, w cannot directly be interpreted in behavioral terms. Most crucially, w differs from the "behavioral Weber fraction" w', as originally defined by Weber. The latter is usually calculated by dividing participants' just-noticeable difference Δn (the amount of stimulus change that can be detected at a fixed level of performance) by the stimulus magnitude n : $w' = \frac{\Delta n}{n}$. Definition of the just-noticeable difference requires specification of an arbitary level of performance (typically 75% correct). Furthermore, the relation between w and w' typically depends on the task. To translate from internal representation to behavior, it is necessary to propose a model of decision making in a given task. In what follows, we adopt the simplest possible decision-making model, which assumes that participants are ideal observers of their internal quantity representation. This assumption allows us to derive simple equations that relate behavioral performance to the internal w.

I.2. Psychophysics of large-smaller judgment

In this task, participants classify a target set of objects with numerosity n as being larger or smaller than a fixed numerical reference. The latter can be specified, as was done in our experiments, by the presentation of several previous sets with a fixed habituation numerosity n_{hab} ("reminder" paradigm as discussed by Mac-Millan and Creelman, 1991).

Suppose that the target numerosities are approximately distributed symmetrically around the criterion numerosity on a logarithmic scale (this is equivalent to saying that there are equal amount of deviations above and below the habituation value, with the amount of deviation defined by the ratio of the two numbers; this was approximately the case in our stimulus set, because we used small deviations on a linear scale). Then the optimal response strategy, under a maximum likelihood criterion, consists in responding "larger" whenever the internal sample coding for the target numerosity falls above some criterion value c. If the participant is unbiased, c should coincide with the internal representation of the reference numerosity, $c = \text{Log}(n_{\text{hab}})$.

Given this response strategy, the probability of responding "larger" to a stimulus numerosity n is given by

$$P_{\text{larger}}(n, n_{\text{hab}}) = \int_{\text{Log}(n_{\text{hab}})}^{+\infty} \frac{e^{-\frac{(r-\text{Log}(n))^2}{2w^2}}}{\sqrt{2\pi}\, w}\, dr$$

which can be simplified into

$$P_{\text{larger}}(n, n_{\text{hab}}) = \int_{0}^{+\infty} \frac{e^{-\frac{(r-\text{Log}(\frac{n}{n_{\text{hab}}}))^2}{2w^2}}}{\sqrt{2\pi}\, w}\, dr = \frac{1}{2}\left(1 + \text{erf}\left(\frac{\text{Log}(\frac{n}{n_{\text{hab}}})}{\sqrt{2}\, w}\right)\right)$$

This is the equation that was fitted to the behavioral data from the larger-smaller task (figure 2E,F,G), with a single free parameter w. Note that performance is predicted to be a symmetrical function of the log ratio of the numerosities n and n_{hab}. Furthermore, in the unbiased case, this function should pass through 0.50 (50% responses or chance level) precisely when $n=n_{\text{hab}}$. Both predictions were verified in most of our analyses. However, a small bias was identified when the data were split into sets of dots versus sets of triangles. Fitting of those data (figure 2G) required a model with two free parameters c and w, letting criterion c differ from its optimal value $c = \text{Log}(n_{\text{hab}})$. The fitting function was then

$$P_{\text{larger}}(n, n_{\text{hab}}) = \int_{c}^{+\infty} \frac{e^{-\frac{(r-\text{Log}(\frac{n}{n_{\text{hab}}}))^2}{2w^2}}}{\sqrt{2\pi}\, w}\, dr$$

I.3. Psychophysics of same-different judgment

In this task, participants classify target sets as having either the same numerosity as the habituation value (e.g. 16), or a different numerosity (e.g. 8, 10, 13, 20, 24, or 32). Again, if the different target numerosities are symmetrically distributed around the habituation numerosity on a logarithmic scale (which was the case in our stimulus set), then the subject's optimal criterion, under a maximum a-posteriori probability rule, consists in responding "same" whenever the internal sample falls within a symmetrical decision interval $[\text{Log}[n_{\text{hab}}] - \delta, \text{Log}[n_{\text{hab}}] + \delta]$ centered on the habituation number n_{hab}, and to respond "different" if the internal sample falls outside this interval.

Given this response strategy, the probability of responding "different" to a stimulus numerosity n is given by

$$P_{\text{different}}(n, n_{\text{hab}}) = \int_{\text{Log}(n_{\text{hab}})-\delta}^{\text{Log}(n_{\text{hab}})+\delta} \frac{e^{-\frac{(r-\text{Log}(n))^2}{2w^2}}}{\sqrt{2\pi}\, w}\, dr$$

which can be simplified into

$$P_{\text{different}}(n, n_{\text{hab}}) = \int_{-\delta}^{+\delta} \frac{e^{-\frac{(r-\text{Log}(\frac{n}{n_{\text{hab}}}))^2}{2w^2}}}{\sqrt{2\pi}\, w}\, dr = \frac{1}{2}\left(\text{erf}\left(\frac{\delta+\text{Log}(\frac{n}{n_{\text{hab}}})}{\sqrt{2}\, w}\right) + \text{erf}\left(\frac{\delta-\text{Log}(\frac{n}{n_{\text{hab}}})}{\sqrt{2}\, w}\right)\right)$$

This is the equation that was fitted to the behavioral data from the change-detection task (figure 2A,B,C), with two free parameters δ and w. Note that performance is predicted to be a symmetrical function of the log ratio of the numerosities n and n_{hab}. Furthermore, if δ is small, this function should be close to a Gaussian. Both predictions were verified experimentally in our data, except when the data were split into sets of dots versus sets of triangles. Fitting of those data (figure 2D) required a model with three free parameters δ, β and w, instantiating an asymmetrical decision interval with a non-zero bias term β. The fitting function was then

$$P_{\text{different}}(n, n_{\text{hab}}) = \int_{\beta-\delta}^{\beta+\delta} \frac{e^{-\frac{(r-\text{Log}(\frac{n}{n_{\text{hab}}}))^2}{2w^2}}}{\sqrt{2\pi}\, w}\, dr$$

II. Neuroimaging

II.1. Neural coding of numerosity

Dehaene and Changeux (1993) postulated on a theoretical basis that numerical quantity could be encoded in the animal and human brain by a population of "numerosity detector" neurons, each tuned to a specific numerosity on an internal logarithmic scale. Nieder and Miller (2002, 2003) provided the first empirical evidence in support of this coding scheme. They identified single neurons tuned to numerosity in the prefrontal cortex and, in smaller proportion, the parietal cortex of awake monkeys that were engaged in a same-different numerosity matching task. Each neuron i had a preferred numerosity p_i, and a bell-shaped tuning curve around this preferred value when responding to any input numerosity n. Nieder and Miller showed that, when numerosity was encoded on a logarithmic scale, the tuning curves could be described most compactly as Gaussian with a fixed width w, identical for all neurons.

Mathematically, then, the firing rate of a given cell i, with preferred numerosity p_i , in response to a stimulus numerosity n, is given by:

$f_i(n) = \alpha_i \, T(n, \, p_i)$,

where the tuning function T_i is the Gaussian $T_i(n) = \frac{e^{-\frac{(\text{Log}[n]-\text{Log}[p_i])^2}{2\,w^2}}}{\sqrt{2\pi}\,w}$,

and α_i is *a* parameter that characterizes the amplitude of the neuron's firing rate.

In this equation, w is the "neural Weber fraction" that defines the degree of coarseness with which neurons encode numerosity. For simplicity, we adopt here the same notation as for the psychophysical Weber fraction defined in paragraph 1. This should not hide that relating neuronal and psychophysical variables remains a difficult task. It typically requires careful analysis of the trial-to-trial variability in firing and of the neuron-to-neuron correlations, and theoretical modeling to determine the number of neurons and pooling rule underlying psychophysical judgments (see Shadlen, Britten, Newsome & Movshon, 1998; Parker and Newsome, 1998). Such work has yet to be undertaken for numerosity judgements.

II.2. Habituation of firing rate

The response of a neuron may vary in time as a function of whether the neuron habituates to the repeated presentation of the same stimulus. A simple hypothesis is that each neuron habituates in direct proportion to how well it is tuned to the current stimulus. Thus, the amplitude of firing rate α_i varies in time according to which numerical stimulus $n(t)$ is presented at time t. A simple, though hypothetical rule for this temporal evolution is:

$\frac{d\,\alpha_i}{d\,t} = -\,\beta\, T_i(n(t)) + \gamma\,(\,\alpha_0 - \alpha_i(t)\,)$

where α_0 characterizes the baseline amplitude of cell firing, β is the rate of habituation, and γ is the rate of recovery from habituation (for simplicity, we assume that on average, α_0 and β do not vary across different neurons with different preferred numerosities p_i, though it should be acknowledged that this has not been verified experimentally).

After a sufficiently long period of habituation with stimulus n_{hab}, solving this equation gives

$\alpha_\infty = \alpha_0 - \frac{\beta}{\gamma}\, T_i(n_{\text{hab}})$

Therefore the firing rate of cell i in response to seeing a novel stimulus n after a period of habituation to stimulus n_{hab} is:

$f_i(n, \, n_{\text{hab}}) = \left(\alpha_0 - \frac{\beta}{\gamma}\, T_i(n_{\text{hab}})\right) T_i(n)$

It it worth stressing several consequences of this equation. First it implies that habituation has a multiplicative effect on neuronal firing: the curve indexing the tuning of the cell to the second stimulus n is unchanged, only the intensity of its firing is modulated. Second, it implies that habituation is a continuous phenomenon : there can be graded levels of adaptation, rather than a categorical distinction between repeated and non-repeated situations. Third, the same tuning function underlies the amount of habituation and the neuron's tuning. In particular, for number neurons, habituation should be measured in logarithmic units of numerical distance. These predictions could be tested in an electrophysiological version of the numerosity habituation experiment with awake monkeys.

II.3. Habituation of brain imaging signals

We assume that the brain imaging signal I during an fMRI priming experiment is linearly related to the total firing rate of the population of neurons in the measured voxel (The possibility that the fMRI signal may reflect synaptic potentials rather than spikes should not fundamentally alter the present mathematical formalism, because in the present case local firing rate and synaptic activity are expected to be highly correlated; for recent discussions, see Logothetis, 2003; Smith et al., 2002).

$I(n, n_{\text{hab}}) = I_0 \sum_{i=1}^{k} f_i(n, n_{\text{hab}})$

Here an assumption is needed about the number of neurons dedicate to a preferred numerosity p. We may follow Dehaene and Changeux (1993) and assume that, in first approximation, neurons form an homogeneous sampling of the logarithmic number line. That is, there are approximately equal numbers of neurons dedicated to each interval of the logarithmic number line (and thus progressively less neurons dedicated to increasingly larger numerosities). This implies that the number of neurons coding for a small interval of preferred numerosities $[p, p + dp)$ is proportional to $\frac{dp}{p}$. In that case, the above sum can be approximated by the integral

$$I(n, n_{\text{hab}}) = I_0 \int_0^{+\infty} \left(\alpha_0 - \frac{\beta}{\gamma} \frac{e^{-\frac{(\text{Log}[n_{\text{hab}}]-\text{Log}[p])^2}{2w^2}}}{\sqrt{2\pi}\, w}\right) \frac{e^{-\frac{(\text{Log}[n]-\text{Log}[p])^2}{2w^2}}}{\sqrt{2\pi}\, w} \frac{dp}{p}$$

which is

$$I(n, n_{\text{hab}}) = \lambda - \mu \frac{e^{-\frac{(\text{Log}[\frac{n}{n_{\text{hab}}}])^2}{4w^2}}}{2\sqrt{\pi}\, w} \quad \text{with } \lambda = I_0 \alpha_0 \text{ and } \mu = I_0 \frac{\beta}{\gamma}$$

Thus, the predicted brain-activation function in numerosity-coding regions is a Gaussian function of the logarithm of the ratio of the two numbers n and n_{hab}. Note that this Gaussian has a standard deviation equal to $\sqrt{2}\,w$, where w is the original standard deviation of the neuronal tuning curves, also plotted on a logarithmic axis. This is the equation that was fitted to the activation curves from our fMRI habituation experiment (figure 4), with three free parameters λ, μ, and w.

References:

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*, 1-42.

Dehaene, S., & Changeux, J. P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience, 5*, 390-407.

Green, D., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Krieger Publishing Company.

Logothetis, N. K. (2003). The underpinnings of the BOLD functional magnetic resonance imaging signal. *J Neurosci, 23*(10), 3963-3971.

MacMillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

Nieder, A., Freedman, D. J., & Miller, E. K. (2002). Representation of the quantity of visual items in the primate prefrontal βcortex. *Science, 297*(5587), 1708-1711.

Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude. Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron, 37*(1), 149-157.

Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: Probing the physiology of perception. *Annu Rev Neurosci, 21*, 227-277.

Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J Neurosci, 16*(4), 1486-1510.

Shepard, R. N., Kilpatrick, D. W., & Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive Psychology, 7*, 82-138.

Smith, A. J., Blumenfeld, H., Behar, K. L., Rothman, D. L., Shulman, R. G., & Hyder, F. (2002). Cerebral energetics and spiking frequency: the neurophysiological basis of fMRI. *Proc Natl Acad Sci U S A, 99*(16), 10765-10770.

van Oeffelen, M. P., & Vos, P. G. (1982). A probabilistic model for the

discrimination of visual number. *Perception & Psychophysics, 32*, 163-170.