

Neuron, Volume 58

Supplemental Data

One-Dimensional Dynamics of Attention and Decision Making in LIP

**Surya Ganguli, James W. Bisley, Jamie D. Roitman, Michael N. Shadlen,
Michael E. Goldberg, and Kenneth D. Miller**

Supplementary Material

Data Analysis

The change of coordinates in firing rate space, shown in both Figures 3C and 3D and Figure 6, of the main paper was performed by first estimating the spontaneous firing rate S_i for each recorded neuron $i=1\dots N$ (where N is now the number of recorded neurons in the given monkey) using 400 msec of recorded data during fixation, prior to the beginning of each trial. The presumed slow mode is then the N -dimensional spontaneous vector \vec{S} whose components are S_i . We normalize \vec{S} to obtain the unit vector $\hat{S} = \vec{S}/|\vec{S}|$. If $\vec{r}(t)$ is the instantaneous firing rate at a given time t , the component along the spontaneous vector \vec{S} (the solid lines in Figures 5C, 5D and 6) is given by $r_s(t) = \vec{r}(t) \cdot \hat{S}$. The norm of the orthogonal component (the dashed lines in the same figures) is given by $|\vec{r} - r_s \hat{S}|$.

The correlation coefficient (the dotted lines) differs from r_s because the mean across neurons is subtracted from the \vec{r} and \vec{S} vectors. Let $\vec{1}$ be the vector whose elements are all equal, normalized so that $\vec{1} \cdot \vec{1} = 1/N$. Then $\vec{r}(t) \cdot \vec{1}$ is simply the instantaneous mean firing rate across neurons at time t . Let $\vec{r}'(t) = \vec{r}(t) - N\vec{1}(\vec{r}(t) \cdot \vec{1})$ and $\vec{S}' = \vec{S} - N\vec{1}(\vec{S} \cdot \vec{1})$. Then the correlation coefficient is $\vec{r}'(t) \cdot \vec{S}' / |\vec{r}'(t)| |\vec{S}'|$.

Linear Model Analysis

In the main paper it is argued that a rapid reduction of LIP rate dynamics to a single dimension is the key to understanding the common crossing time of all the LIP neurons. The purpose of the remaining sections is to show concretely that this rapid reduction can be implemented generically in a linear model of the rate dynamics, without any need for fine tuning. We choose a linear model because it is the simplest model that can fit the observed firing rate data. Also the linear model has the advantage that it can be analyzed exactly, yielding intuition for how the common crossing time can arise robustly. However we stress that the underlying idea of rapid dimensionality reduction can be equally well implemented in more complex, nonlinear models and we are not committed to a linear model of actual LIP dynamics.

Single Neuron Dynamics and the Fine Tuning Problem.

We begin by reviewing the linear rate dynamics (see also Seung, 2003) of a single neuron with time constant τ , a recurrent excitatory autapse w , and external input I . The firing rate $r(t)$ obeys the differential equation

$$\tau \frac{dr}{dt} = -r + wr + I. \quad (1)$$

Here $w < 1$ to ensure the stability of (1). We can rewrite this equation as

$$\frac{\tau}{1-w} \frac{dr}{dt} = -r + \frac{I}{1-w}. \quad (2)$$

We see two important effects due to the recurrent feedback w . First, the effective input is amplified

by the factor $\frac{1}{1-w}$, leading to an amplified steady state response $\frac{I}{1-w}$ obtained by setting the derivative on the left hand side of (2) to zero. Second, the effective time constant $\tilde{\tau}$ multiplying the derivative,

$$\tilde{\tau} = \frac{\tau}{1-w}, \quad (3)$$

is also amplified by the same factor, leading to a slower dynamics. This slowness is essentially the price that is paid for amplification via recurrent feedback.

In this simple model, we wish to explore the dependence of the peak visual response V , the delay period activity D , and the effective time constant $\tilde{\tau}$ on the various parameters. We model the visual response by turning on a constant visual input I^V for $t_o = 100$ msec, as is done in the experiment. Assuming the initial rate $r(0)$ is negligible, the time course of the visual response while the input is present, obtained by solving (2), is

$$r(t) = (1 - e^{-t/\tilde{\tau}}) \frac{I^V}{1-w} \quad \text{for } 0 \leq t \leq t_o \quad (4)$$

Again $\tilde{\tau}$ is the effective time constant given in (3), not the single neuron time constant τ . The peak visual response is then obtained by setting $t = t_o$ in (4),

$$V = (1 - e^{-t_o/\tilde{\tau}}) \frac{I^V}{1-w}. \quad (5)$$

We next model the delay period activity D as the steady state response to a top-down feedback input I^T . This yields immediately

$$D = \frac{I^T}{1-w}. \quad (6)$$

So now equations (3), (5) and (6) give us the relationship between the observed quantities $\tilde{\tau}$, V and D in terms of the model parameters τ , w , I^T and I^V .

Now suppose we have a collection of neurons in LIP with various values for the 4 model parameters. The heterogeneity of the neural data forces us to choose diverse values of the model parameters for each neuron. However, the experimental data provides a strong constraint that for every neuron i ,

$$\tilde{\tau}_i \ln \frac{V_i}{D_i} \approx t_c, \quad (7)$$

where t_c is the common crossing time of all the neurons. Using Eqs. (3),(5)-(6), this becomes

$$\tilde{\tau}_i \ln \left[\left(1 - e^{-t_o/\tilde{\tau}_i} \right) \frac{I_i^V}{I_i^T} \right] \approx t_c. \quad (8)$$

In the regime $t_o/\tilde{\tau}_i \ll 1$, when neuronal decay times are longer than the input stimulus duration, as is our case, this can be rewritten as

$$\tilde{\tau}_i \ln \frac{t_o I_i^V}{\tilde{\tau}_i I_i^T} \approx t_c. \quad (9)$$

Thus, to realize this experimental constraint, one must impose biophysically unreasonable constraints on the diverse model parameters for each neuron. For example, in (9), suppose a neuron has a particularly long decay time $\tilde{\tau}_i$. The left hand side is dominated by a linear increase in $\tilde{\tau}_i$

which means the ratio of inputs $\frac{I_i^V}{I_i^T}$ inside the logarithm will have to decrease exponentially to compensate for the larger $\tilde{\tau}_i$ in order to maintain the constraint. Similarly suppose a neuron

receives a particularly large visual input I_i^V , but a small top down feedback signal I_i^T . Such a neuron must have its intrinsic neuronal time constant τ_i and recurrent excitation w_i tuned so that its observed decay time $\tilde{\tau}_i$ is small enough to compensate for the large ratio of inputs.

More generally, there is no biophysical basis for postulating various unrealistic constraints, of the type above, that relate the top-down feedback input, bottom-up visual input, strength of connectivity and neuronal time constants to each other. Hence we say that the experimental data cannot be explained at the single neuron level without severe fine tuning problems, at least in this simple model. A cursory examination of more complex nonlinear models (unpublished) indicates that any such single neuron explanation will likely suffer from a similar fine tuning problem.

A generic solution through linear network dynamics.

We now explore how recurrent network connectivity can realize the observed constraint (7) without fine tuning. Following the main paper, first consider N neurons sharing the same receptive field. Their dynamics is governed by the vector equation

$$\tau \frac{d\vec{r}}{dt} = -\vec{r} + W\vec{r} + \vec{I}. \quad (10)$$

Here \vec{r} is an N dimensional vector whose i 'th component is the firing rate of neuron i . W is an N by N matrix whose elements w_{ij} represent the strength of recurrent excitation from neuron j to neuron i . For simplicity, we choose each neuron i to have the same time constant τ . Alternatively we could, and later will, replace τ with a diagonal matrix T of different time constants for each neuron. The initial choice of a single time constant yields a simpler conceptual understanding of how the common crossing time arises. As in the first section, we will consider two different inputs, a visual input \vec{I}^V and a top-down input \vec{I}^T which give rise to the visual transient \vec{V} and the delay period activity \vec{D} respectively.

In order to analyze (10), it is useful to perform a change of basis to the right eigenvectors of the matrix W . One can find a change of basis matrix U and a diagonal matrix Λ such that

$$U^{-1}WU = \Lambda. \quad (11)$$

The elements λ_i along the diagonal of Λ are the eigenvalues of W . If one performs a change of basis by substituting $\vec{r} = U\vec{c}$ and $\vec{I} = U\vec{\tilde{I}}$ into (10), then the dynamics decouples in terms of the new variables \vec{c} . Each component c_i of \vec{c} obeys an independent equation

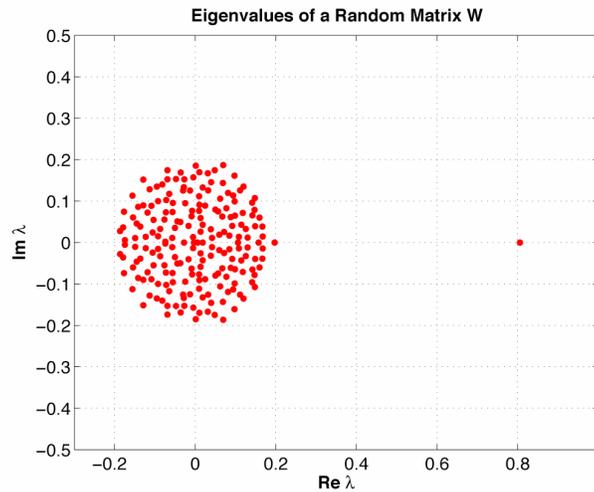
$$\frac{\tau}{1-\lambda_i} \frac{dc_i}{dt} = -c_i + \frac{\tilde{I}_i}{1-\lambda_i}. \quad (12)$$

One can think of the dynamical variables $c_i(t)$ as the N "eigenmodes" of the N neurons. These modes evolve independently of each other. Here \tilde{I}_i is the input to the i 'th eigenmode. \tilde{I}_i in the new mode basis is related to I_j in the old neuron basis by $\tilde{I}_i = \sum_{j=1}^N U_{ij}^{-1} I_j$. At any given time, the activities of the neurons can be recovered from the activities of the modes using $r_i = \sum_{j=1}^N U_{ij} c_j$.

The time evolution of the eigenmodes c_i in (12) is exactly equivalent to that of a single neuron in (2) where the eigenvalue λ_i of the mode plays the role of the recurrent connectivity w . Thus all the intuition gained in the previous section applies to the dynamics of the modes. Modes whose eigenvalues are close to one amplify their inputs by a factor $\frac{1}{1-\lambda_i}$ but are very slow at reaching their steady state values. On the other hand, modes whose eigenvalues are closer to zero do not amplify their inputs much, but are very fast and rapidly approach equilibrium. If we try to

explain the common crossing time t_c via a rapid reduction in the dynamics to a single dimension, then it is now clear that we can arrange for this rapid reduction by having precisely one eigenvalue of W close to one, while the rest of the eigenvalues are close to zero.

We now show that we can achieve this situation for very generic choices of the synaptic connectivity matrix W , without relying on any fine tuning. In fact this situation holds if we choose the elements w_{ij} essentially at random. In keeping with anatomical data showing that cortical connectivity is sparse, with rarely more than a 10% chance of any two randomly chosen neurons sharing a synaptic connection (Holmgren et al., 2003; Markram, et al., 1997), we will choose each w_{ij} to be nonzero with probability $p=0.1$. When w_{ij} is chosen to be nonzero, we will set $w_{ij} = w/N$, where w is drawn for each i and j from a gaussian distribution with mean μ_w and variance σ_w^2 . Thus W is a random N by N matrix whose entries are drawn independently from a distribution described by three parameters, p , μ_w , and σ_w^2 . Supplementary Figure 1 shows the eigenvalue distribution for a random matrix drawn from this distribution with parameters $N = 200$, $p = 0.1$, $\mu_w = 8$, and $\sigma_w^2 = 4$. We see that the spectrum is exactly what we desired. There is a circular cloud of eigenvalues near the origin of complex plane, and precisely one eigenvalue far out on the real axis, close to 1.



Supplementary Figure 1.

One can heuristically understand the structure of this eigenvalue spectrum, and its dependence on the parameters, as follows. We can write $W = M + J$ where M is the mean matrix, in which every element M_{ij} assumes the same value $\frac{1}{N} p \mu_w$, the mean value of the elements of W . J is then a random matrix whose elements have zero mean, and variance.

$$\sigma^2 = (p\sigma_w^2 + \mu_w^2 p(1-p))/N^2$$

The eigenvalues of M are easily understood: the uniform vector, consisting of all neurons firing at the same rate, is an eigenvector of M with eigenvalue $p\mu_w$. Every other vector orthogonal to this uniform mode consists of elements that sum to zero and hence is an eigenvector of M with eigenvalue zero. That is, a matrix with uniform elements naturally has a gap in its set of eigenvalues, as we are looking for, with one positive eigenvalue and all other eigenvalues equal to zero.

The eigenvalues of J are also well understood from results in random matrix theory. Girko's

circular law (Girko, 1984) states that an N by N random matrix, with entries drawn independently from a gaussian distribution with zero mean and fixed variance σ^2 , has eigenvalues that are uniformly distributed within a circle in the complex plane centered at the origin. The radius of this circle is $\sigma\sqrt{N}$. Although the elements of J do not come from a Gaussian distribution, the circular law depends essentially on the standard deviation of the distribution and is a good approximation for large enough N . Thus, the radius of the cloud does not depend in detail on the the matrix elements of J , and is given by

$$R = \sqrt{\frac{p\sigma_w^2 + \mu_w^2 p(1-p)}{N}}. \quad (13)$$

With the above parameter values, the radius R of the eigenvalue cloud, given theoretically in (13), is 0.19, which agrees with the picture in Supplementary Figure 1.

The most important feature of (13) is the $\frac{1}{\sqrt{N}}$ dependence of R . We see that by increasing N , we can compress the cloud of eigenvalues near the origin. Compressing the eigenvalues along the real axis makes the associated eigenmodes decay even faster. Compressing them along the imaginary axis suppresses any oscillations.

The full distribution of eigenvalues of W can be understood at least heuristically as a perturbation of the spectrum of M induced by J . We have checked numerically that a single eigenvalue of the full random matrix W always lies with high probability close to the single nonzero eigenvalue $p\mu_w$ of M . In fact the standard deviation of its position is equal to the standard deviation σ of the weights of W , without the extra factor of \sqrt{N} that contributes to the radius of the cloud. This can be proven using matrix perturbation theory, or checked numerically. In any case, heuristically we do not expect this eigenvalue to be much perturbed because J acting on the uniform vector gives a vector whose elements have zero mean and variance $\propto N\sigma^2 \propto 1/N$, meaning that the uniform vector is nearly an eigenvector of J with eigenvalue 0. The cloud of other eigenvalues near the origin represents a breaking of the degeneracy of the zero eigenvalues of M , as these eigenvalues spread out into a cloud whose radius is determined by the variance of the elements of J .

It is important to note that while the radius of this cloud scales as $\frac{1}{\sqrt{N}}$, the location of the maximal eigenvalue is independent of N , implying that the time scale associated with this slow mode is independent of the number of neurons in the modelled patch. This independence of the slowest time scale on the number of neurons crucially depends on our $1/N$ scaling of the weights, which was necessary for the network to remain stable (real part of maximal eigenvalue < 1) as the number of neurons in it, N , was scaled up. Once weights are fixed, however, involving more neurons in a pattern increases the overall recurrent excitation involved and leads to a slower decay of that pattern. In the nonlinear scenario, one can actually fix the weights, since stability can be achieved by nonlinear saturating effects, rather than weight rescaling as the number of neurons involved in any pattern increases. Thus in the nonlinear case, modulation of the number of participating neurons in any given neuronal activity pattern can act as powerful method to control the decay time scale of that pattern.

In Rajan and Abbott, 2006, it was shown that the same analysis will apply to a more realistic weight matrix that results when one considers that there are separate populations of excitatory and inhibitory neurons. Again, so long as the mean over all weights is positive -- that is, excitation and inhibition are not quite balanced, but instead each cell receives a slight excess of excitation -- there will be one eigenvalue corresponding to this nonzero mean, and a cloud of eigenvalues near the origin induced by the random variations in weights.

Armed with these results we can now connect to the data. To do so, we must first choose numerical values for the various parameters to fit the observed data. We have already chosen the sparse connectivity $p = 0.1$ in accordance with anatomical data. Our next choice is the time scale τ for the rate dynamics. It is thought (Shriki, et al., 2003) that in the reduction from biophysical spiking networks to phenomenological rate models, the dominant contribution to the time scale τ comes from the time scale at which NMDA receptors inactivate. This time scale can range from 40-80 msec (Lester, et al., 1990). We simply choose τ to be 60 msec. However the observed decay time constants of the neurons recorded in Bisley and Goldberg, 2005, are much longer, and are on the order of 300 msec. This long time scale most likely occurs due to a slow mode arising out of recurrent excitation. We have seen that the maximal eigenvalue of the random connectivity matrix W is very close to $p\mu_w$, yielding a longest network time scale $\frac{\tau}{1-p\mu_w}$. Matching this longest time scale to 300 msec allows us to solve for μ_w , yielding $\mu_w = 8$. Furthermore we allow for heterogeneity in the synaptic weights by choosing $\sigma_w = 4$. This parameter is not crucial, since substantial heterogeneity is already contributed by p .

We can also tolerate further heterogeneity by allowing the actual time constants τ_i of each neuron to fluctuate about their mean value τ . In keeping with the variance in NMDA receptor closing times, we choose τ_i from a gaussian distribution with mean τ already chosen to be 60 msec, and a standard deviation $\sigma_\tau = 20$ msec. The relevant random matrix is no longer W but rather $(I-W)^{-1}T$, where T is the matrix of time constants. However the essential intuition presented above remains the same for this matrix. If the matrix assumes its mean value, it will consist of one eigenvalue equal to $\frac{\tau}{1-p\mu_w}$ and the rest of the eigenvalues equal to τ . Just like in the case of W , the degeneracy of eigenvalues in the random case expands into a cloud centered at $\tau = 60$ msec, while the outlier eigenvalue remains at 300 msec. The essential structure of one slow mode and multiple fast modes is preserved.

With the mean values τ , p and μ_w fixed, we must choose the inputs to each neuron \bar{I}_i^V and \bar{I}_i^T . These two inputs independently determine the observed range of peak visual responses V_i and delay activities D_i . In keeping with the idea that no individual parameters in the linear model (10) have to be fine tuned, just as we picked the connectivity matrix randomly, we also pick the visual and top-down inputs randomly. In fact we draw them from a uniform distribution between a minimum and maximum firing rate. This uniform distribution is the maximal entropy distribution consistent with the fact that neurons have limits on their firing rate, and hence this choice introduces no further assumptions about the inputs. In order to qualitatively fit the range of peak visual responses, we choose $I_{min}^V = 80$ Hz and $I_{max}^V = 200$ Hz, and to fit the range of delay period activities, we choose $I_{min}^T = 10$ Hz and $I_{max}^T = 30$ Hz. These values are consistent, in the former case, with the firing rates of visually responsive neurons in lower areas thought to provide feedforward inputs to LIP, and, in the latter case, with the firing rates of neurons in the frontal eye field thought to participate in a top-down feedback signal to LIP.

With these natural parameter choices, we have seen in the main paper that the constraint across neurons (7) arises essentially for free, as long as N is large enough. Again, the essential idea is that since the largest eigenvalue of W is close to $p\mu_w = 0.8$, whereas the eigenvalue with the next largest real part scales as $1/\sqrt{N}$ in (13), for large enough N there is a separation of time

scales between the slow mode and all the other modes. After the stimulus turns off, all the fast modes decay quickly with a time constant close to $\tau = 60$ msec, leaving only the slow mode at longer time scales. When the neurons cross their delay activity, the dynamics has already become one dimensional, yielding the common crossing time for all neurons in the same patch, as explained in the main paper.

Also discussed in the main paper was the correlation between spontaneous activity, and peak visual as well as delay activity. For $N=100$ neurons it was found that there exists a correlation between spontaneous and delay activity, but not between spontaneous and visual. In this simple linear model, such correlations depend slightly on N . As N increases, there are more and more fast modes (precisely $N-1$ of them), and their presence contributes to a slight decay of the correlation coefficient between spontaneous and both delay as well as visual transients. However for any N , spontaneous is more correlated with delay than with visual, as discussed in the main body. Furthermore, for N in the range of 100 to 1000, which may be roughly the number of strong synapses activated by a single input to a visual cortical neuron, the observed correlations are consistent with the data.

Invariance of the crossing time across different patches.

We have analyzed how a common crossing time can be achieved generically in a single patch, but we must also explain why the crossing time of neurons, chosen in different patches that are not coupled to each other, nevertheless remains invariant across such neurons. We can think of the crossing time, expressed as the constant of proportionality in (7), as a random variable itself. Our goal is to obtain a rough estimate of its variance.

Since the time at which the neurons cross their delay activity is dominated by the slow one dimensional dynamics of a single mode $c_s(t)$, the dominant contribution to the crossing time t_c arises from the time the slow mode $c_s(t)$ in the visual transient crosses its own level of excitation in the delay period. We can ignore the contribution of all the faster modes because of their rapid decay. The crossing time of the slow mode itself depends on three quantities: the slow mode's peak excitation level V_s due to visual input, its delay period excitation level D_s , and its decay time constant τ_s . We have already seen from the random matrix results that the time constant τ_s is,

with high probability, close to $\frac{\tau}{1 - p\mu_w}$. Now the peak slow mode visual response V_s depends on

the total input to that mode. The input to this mode is given by the inner product of the randomly chosen neuronal input vector \vec{I}^V with the eigenvector associated with the slow mode. This eigenvector is random since the connectivity matrix W is random, but as we have seen above, it is a uniform mode with all positive entries. Thus the visual input to the slow mode, and hence the peak visual response V_s , arises from the sum of N independent random numbers; therefore it must have coefficient of variation (standard deviation over mean) that scales as $1/\sqrt{N}$. A similar logic applies to D_s .

So for randomly chosen connectivities W and inputs \vec{I}^V and \vec{I}^D , the properties of the slow mode, τ_s , V_s and D_s are strongly peaked about their mean values. Exploiting this fact, we calculate the dependence of the mean value of t_c on the parameters of the various distributions we have chosen. In the equations below, each of the quantities τ_s , V_s , D_s as well as t_c are to be thought of as the mean values of the associated variables which arise in a large random network with random inputs. We can treat the slow mode as if it were a single neuron, yielding

$$\begin{aligned}
V_s &\propto (1 - e^{-t_0/\tau_s}) \frac{I_{mean}^V}{1 - p\mu_w} \\
D_s &\propto \frac{I_{mean}^T}{1 - p\mu_w}.
\end{aligned}
\tag{14}$$

Now solving for t_c in $V_s e^{-t_c/\tau_s} = D_s$, yields the final dependence of t_c on parameters,

$$t_c = \frac{\tau_s}{\ln\left[(1 - e^{-t_0/\tau_s}) \frac{I_{mean}^V}{I_{mean}^T}\right]}.
\tag{15}$$

The final result is intuitive: the time scale of the slow mode τ_s sets the time scale of the crossing time, while the ratio of feedforward visual input to top-down feedback signal can tune this crossing time, essentially setting the level of distractibility of the monkey. Due to $1/\sqrt{N}$ averaging effects, for each randomly chosen patch, the probability distribution for the crossing time t_c of that patch is strongly peaked about its mean value in (15). Hence t_c , which controls the distractibility of the monkey, is an emergent property that does not depend on the details of each patch, but rather on the gross statistical properties of the distribution from which the patches are drawn.

References

- Bisley, J.W., and Goldberg, M.E. (2005). Neural correlates of attention and distractibility in the lateral intraparietal area. *J Neurophys* 95, 1696–1717.
- Girko, V.L. (1984). Circular law. *Theory Probab Appl* 29, 694–706.
- Holmgren, C., Harkany, T., Svennenfors, B., and Zilberter, Y. (2003). Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. *J Physiol* 551, 139–153.
- Lester, R.A., Clements, J., Westbrook, G., and Jahr, C.E. (1990). Channel kinetics determine the time course of NMDA receptor-mediated synaptic currents. *Nature* 346, 565–567.
- Markram, H., Lubke, J., Frotscher, M., Roth, A., and Sakmann, B. (1997). Physiology and anatomy of synaptic connections between thick tufted pyramidal neurons in developing rat neocortex. *J Physiol* 15, 409–440.
- Rajan, K., and Abbott, L. (2006). Eigenvalue spectra of random matrices for neural networks. *Phys Rev Lett* 97, 188104.
- Seung, H.S. (2003). *Handbook of Brain Theory and Neural Networks*, chap. Amplification, Attenuation, and Integration, pp. 94–97. (Cambridge, MA: MIT Press).
- Shriki, O., Hansel, D., and Sompolinsky, H. (2003). Rate models for conductance-based cortical neuronal networks. *Neural Comp* 8, 1809–1841.