Technical note

# Simulating brain signals with predefined mutual correlations

Alexander Moiseev[a,*]

[a] *Behavioral and Cognitive Neuroscience Institute, Simon Fraser University, 8888 University Ave, Burnaby, BC V5A 1S6, Canada*

**Abstract**

This technical note describes a simple linear algorithm which allows generating arbitrary number of signal time courses whose mutual correlations are defined by a given mathematically consistent correlation matrix. This is achieved by mixing some seed set of waveforms typically reflecting desired specific features of the target brain signals. If needed the resulting signals can be properly tapered to be neurophysiologically plausible. For multi-epoch designs the generated waveforms may also include a "phase-locked" component which does not vary between the epochs.

*Keywords:* M/EEG, fMRI, brain signals simulation, correlated brain signals

## 1. Introduction

In the field of functional brain imaging using M/EEG or fMRI one often needs to validate various techniques by means of computer simulations. Sim-

*Corresponding author

*Email address:* amoiseev@sfu.ca (Alexander Moiseev)

ulating the human brain in general is a vast area of research [1], but in this technical note we only focus on a small but commonly encountered task of simulating sets of brain signals whose time courses have predefined mutual correlations. More often than not we also want these signals to resemble those observed in practice. For multi-epoch designs we may want the signal time courses to consist of a part which is identical for all epochs to represent the "evoked" or "phase-locked" component of the brain activity, and a random part changing from epoch to epoch (the oscillatory, or "time-locked" component).

For just a single pair of sources the desired correlation may be easily obtained, for example, by using sinusoidal signals with a properly selected phase difference. However such sinusoidal signals hardly resemble brain activity observed in real life. Simulating larger number of sources with complicated mutual inter-dependencies becomes more challenging. This technical note describes a simple linear mixing algorithm to generate an arbitrary number of source time courses that are mutually correlated with any given mathematically consistent correlation matrix. Specific situation-dependent features of those signals can be controlled by properly choosing an initial "seed" set of waveforms. In some cases a choice or construction of such seed set may not be a trivial task, however the latter lies out of the scope of this short note. Finally, the output waveforms can be properly tapered and include "evoked" parts persisting between the epochs. In this case the required mutual correlations will be obtained for the ultimate composite signals.

## 2. Methods

### *2.1. Problem formulation*

Assume that we have to construct multi-epoch time courses for $n$ sources. Then every epoch is a set of $n$ discreet real valued signals $p_i(t),\ i = 1, ...n$, each having $m$ time points $t$. We require $m >> n$ for the time averaging operations to be statistically robust. Taken together signals $p_i(t)$ can be represented as a $n \times m$ matrix $\boldsymbol{p}$ with $n$ rows (the "source" dimension) and $m$ columns (the "time" dimension). Signals have predefined "evoked" parts $c_i(t)$ that do not change from epoch to epoch, and "time-locked" or oscillatory parts $z_i(t)$ that may differ between epochs. Thus we have

$$\boldsymbol{p} = \boldsymbol{z} + \boldsymbol{c} \tag{1}$$

where $\boldsymbol{z},\ \boldsymbol{c}$ are $n \times m$ matrices.

The following notation will be used further on. Consider an arbitrary set of signals represented by its $n \times m$ matrix $\boldsymbol{a}$. Let source and time domain vectors be $n$-dimensional column and $m$-dimensional row vectors, respectively. Let angle brackets $\langle . \rangle$ denote averaging over time – that is over $m$ time points. Then $\langle a_i(t) \rangle = (1/m)\Sigma_{j=1}^{m} a_{ij}$. In matrix form we have $\langle \boldsymbol{a} \rangle = (1/m)\boldsymbol{a}\boldsymbol{1}_m^T$, where $\boldsymbol{1}_m$ is a $m$-dimensional row vector of ones and superscript "$T$" denotes matrix transposition. Signal $\boldsymbol{a}$ can be divided into its mean $\bar{\boldsymbol{a}}$ and a fluctuating part $\tilde{\boldsymbol{a}}$: $\boldsymbol{a} = \bar{\boldsymbol{a}} + \tilde{\boldsymbol{a}}$; $\bar{\boldsymbol{a}} = \langle \boldsymbol{a} \rangle \boldsymbol{1}_m$; $\tilde{\boldsymbol{a}} = \boldsymbol{a} - \bar{\boldsymbol{a}}$. Finally, let $\boldsymbol{C}_a$ denote a $n \times n$ covariance matrix of the signal set $\boldsymbol{a}$, that is $\boldsymbol{C}_a = \langle \tilde{\boldsymbol{a}}\tilde{\boldsymbol{a}}^T \rangle$. As matrix multiplication in the brackets already involves summation over the time points, we simply set $\boldsymbol{C}_a = \tilde{\boldsymbol{a}}\tilde{\boldsymbol{a}}^T/(m-1)$ (an unbiased estimate of the covariance). Alternatively $1/m$ (maximum likelihood) normalization can be used;

this will not change the final expressions provided the same normalization is applied everywhere. A specific case of a covariance matrix with all diagonal elements equal to 1 is a correlation matrix; to distinguish it we'll write $\hat{\boldsymbol{C}}_a$ instead of $\boldsymbol{C}_a$.

Assume without loss of generality that evoked signals $c_i$ are normalized to have a unit variance: $\langle \tilde{c}_i^2 \rangle = 1$. We want oscillatory parts $z_i,\ i = 1, ..., n$ to be zero mean, noise-like signals that differ between epochs but have the same covariance. Additionally we want $z_i(t)$ to be smoothly tapered at the ends to be physiologically plausible: $z_i(t) = e(t) r_i(t)$, where $r_i(t)$ is some not tapered random signal and the taper function $e(t) = \{e_j\},\ j = 1, ..., m$ is assumed to be the same for all $i = 1, ..., n$ for simplicity. In matrix form this is written as

$$\boldsymbol{z} = \boldsymbol{r}\boldsymbol{E};\quad \langle \boldsymbol{r}\boldsymbol{E} \rangle = 0;\quad \boldsymbol{E} = diag(\boldsymbol{e}) = \{e_i \delta_{ij}\},\ i, j = 1, ..., m \tag{2}$$

where $\delta_{ij}$ is the Kronecker's delta.

To be able to control relative contributions of phase-locked and time-locked components to the full signal, define quotients $q_i,\ 0 \leq q_i \leq 1,\ i = 1, ..., n$ such that $q_i^2$ specifies what part of the variance of the full signal $\langle \tilde{p}_i^2(t) \rangle$ is due to the variance of the evoked part $\langle \tilde{c}_i^2(t) \rangle$, and also request that oscillatory parts $\boldsymbol{z}$ are uncorrelated with $\boldsymbol{c}$: $\langle \boldsymbol{z}\boldsymbol{c}^T \rangle = 0$. Then each $p_i(t)$ can be written as $p_i(t) = z_i(t) + q_i c_i(t)$. Summarizing all of the above and requiring without loss of generality that the target signals $p_i$ have a unit variance: $\langle \tilde{p}_i^2(t) \rangle = 1$, we can now write instead of (1):

$$\boldsymbol{p} = \boldsymbol{r}\boldsymbol{E} + \boldsymbol{Q}\boldsymbol{c};\quad \boldsymbol{Q} = diag(\boldsymbol{q}) = \{q_i \delta_{ij}\},\ i, j = 1, ..., n \tag{3}$$

where

$$\langle \boldsymbol{r}\boldsymbol{E}\rangle = 0;\ \langle \boldsymbol{r}\boldsymbol{E}\tilde{\boldsymbol{c}}^T\rangle = 0;\ \langle \tilde{p}_i^2\rangle = \langle \tilde{c}_i^2\rangle = 1;\ i = 1, ..., n \tag{4}$$

Thus mathematically the problem of generating a set of signals with predefined correlations can be formulated as follows.

PROBLEM. Given unit variance evoked signals $c_i$, quotients $q_i$, $0 \leq q_i \leq 1$, $i = 1, ..., n$ and a taper function $e(t)$, generate a set of random signals $\boldsymbol{r}$ such that simulated "brain signals" $\boldsymbol{p}$ in (3), (4) have a desired correlation matrix $\hat{\boldsymbol{C}}_p$.

*2.2. Solution*

Start with an arbitrary $(n \times m)$ *seed signal set* $\boldsymbol{s}$ which usually should reflect desired features of the modeled brain activity. In the simplest case such a set can be produced, for example, by a random number generator or by filtering white noise to desired frequency bands. For more realistic designs characteristic samples of actually recorded brain activity could be used.

Then for any instance (epoch) of the seed signal set $\boldsymbol{s}$ the required signal set $\boldsymbol{r}$ is obtained by the formula:

$$\boldsymbol{r} = (\hat{\boldsymbol{C}}_p - \boldsymbol{Q}\hat{\boldsymbol{C}}_c\boldsymbol{Q})^{1/2}\boldsymbol{C}_{sPE}^{-1/2}\boldsymbol{s}\boldsymbol{P} \tag{5}$$

and the full signal $\boldsymbol{p}$ is given by Eq. (3). In Eq. (5) $\hat{\boldsymbol{C}}_c$ is the correlation matrix of the evoked parts of the target signal, $\hat{\boldsymbol{C}}_c = \langle \tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}^T\rangle$; $\boldsymbol{P}$ is a time domain $(m \times m)$ projection matrix given by the expression:

$$\boldsymbol{P} = \boldsymbol{I}_m - \boldsymbol{A}\boldsymbol{A}^\dagger;\quad \boldsymbol{A} = \{\boldsymbol{e}^T, \boldsymbol{E}\tilde{\boldsymbol{c}}^T\} \tag{6}$$

where $\boldsymbol{I}_m$ is a $m$-dimensional identity operator; $m \times (n+1)$ rectangular matrix $\boldsymbol{A}$ is obtained by concatenating $\boldsymbol{e}^T$ and columns of the product $\boldsymbol{E}\tilde{\boldsymbol{c}}^T$; symbol "†" denotes matrix pseudo-inverse. Finally, $\boldsymbol{C}_{sPE}$ is a covariance of a projected and tapered seed set $\boldsymbol{s}$:

$$\boldsymbol{C}_{sPE} = \langle \boldsymbol{s}\boldsymbol{P}\boldsymbol{E}^2\boldsymbol{P}\boldsymbol{s}^T \rangle \tag{7}$$

***Remark.*** *With straightforward modifications, solution (5) – (7) can be generalized to the case of complex-valued signals and multiple taper functions* $e_i(t),\ i = 1, ..., n$.

Note that solution (5) only exists if $\hat{\boldsymbol{C}}_p - \boldsymbol{Q}\hat{\boldsymbol{C}}_c\boldsymbol{Q}$ is a positively defined matrix. Indeed, correlations between the evoked signals $c_i(t)$, which are reflected by the term $\boldsymbol{Q}\hat{\boldsymbol{C}}_c\boldsymbol{Q}$, impose a lower limit on correlations between the full signals $p_i(t)$. For example, there always exists large enough $\boldsymbol{Q}$ such that it becomes impossible to create a completely uncorrelated set $\boldsymbol{p}$ (i.e. $\hat{\boldsymbol{C}}_p = \boldsymbol{I}_n$) if evoked parts are at least partially correlated.

A step by step computational procedure implementing solution (5) is outlined in Listing 1. Mathematical proofs are given in *Appendix A*.

Listing 1: Computation steps

**Input**:

$\hat{\boldsymbol{C}}_p, \boldsymbol{c}, \boldsymbol{e}, q_i,\ i = 1, ..., n$

Seed signals $\boldsymbol{s}$ generator function

**Preparation**:

Calculate $\bar{\boldsymbol{c}}, \tilde{\boldsymbol{c}}$

Normalize $c_i(t)$ to get $\langle \tilde{c}_i^2 \rangle = 1,\ i = 1, ..., n$

Calculate $\hat{\boldsymbol{C}}_c = \langle \tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}^T \rangle$

Construct diagonal matrices $\boldsymbol{Q}, \boldsymbol{E}$

**Validation**:

```
    Verify that Ĉ_p is positive definite
    Verify that Ĉ_p − QĈ_cQ is positive definite
Calculation:
    Construct matrix A and projector P using (6)
    Calculate matrix (Ĉ_p − QĈ_cQ)^{1/2}
    for each epoch:
        Generate the seed set s
        Calculate covariance C_sPE = ⟨sPE²Ps^T⟩
        Calculate (C_sPE^{-1})^{1/2}
        Calculate r using (5)
        Calculate p using (3)
    end for
end
```

## 3. Example

Imagine a situation where we want to model an event-related response from three pairs of functionally related brain locations in three main frequency bands. Let sources $\{1, 2\}$ represent alpha band activity in 8 to 12 Hz range, sources $\{3, 4\}$ represent beta activity in 16 to 20 Hz range, and sources $\{5, 6\}$ model gamma activity in 35 to 40 Hz range.

Assume also that alpha and beta signals carry evoked components, while gamma signals don't. The evoked components do not need to belong to corresponding frequency bands. For simplicity let the evoked parts of the alpha pair $\{1, 2\}$ have the same waveforms for both sources but opposite signs, while for the beta pair $\{3, 4\}$ they are also identical and have the same signs. We want the evoked parts to contribute 60% to the variance of the total generated signals: $q_i^2 = 0.6,\ i = 1, ..., 4$.

We would like the full (that is, the oscillatory and evoked parts combined) signals to be correlated, to reflect functional relations between corresponding brain areas. Specifically, we request the following correlation coefficients between the signals of each pair: $-0.9$ for alpha, 0.5 for beta and 0.2 for gamma. Further assume that across the pairs alpha and beta signals are uncorrelated, and so are the beta and gamma signals. However we want a moderate correlation of 0.33 within odd-numbered signals and within even-numbered signals of the alpha and gamma pairs - that is between signals $1, 5$ and signals $2, 6$. The above settings result in the target correlation matrix $\hat{\boldsymbol{C}}_p$ shown in Eq.(8).

$$\hat{\boldsymbol{C}}_p = \begin{pmatrix} 1. & -0.9 & 0. & 0. & 0.33 & 0. \\ -0.9 & 1. & 0. & 0. & 0. & 0.33 \\ 0. & 0. & 1. & 0.5 & 0. & 0. \\ 0. & 0. & 0.5 & 1. & 0. & 0. \\ 0.33 & 0. & 0. & 0. & 1. & 0.2 \\ 0. & 0.33 & 0. & 0. & 0.2 & 1. \end{pmatrix} \tag{8}$$

Finally, we use a taper $e(t) = h(t, T)e^{-6t/T}$ where $h(t, T)$ is a Hanning window function and $T$ is a signal duration; set $T = 1\ s$ for this example. The exponential multiplier is added to model a decay of the event-related signals with time.

Instances of various signals comprising a single epoch of the seed, evoked and target sets which were generated in accordance with the above requirements are depicted in Fig. 1; the Python code for this example can be found at https://github.com/aa-moiseev/pub-corr-src-set.git.

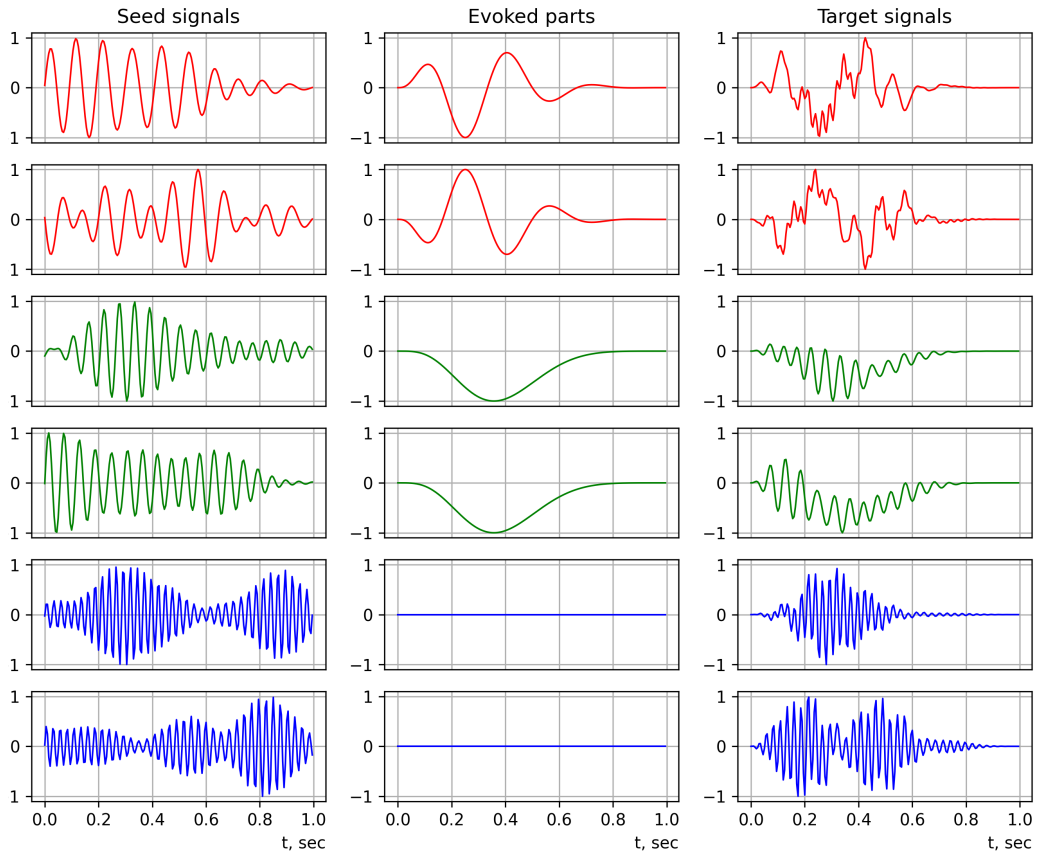


Figure 1: **Seed ($s$), evoked ($c$) and target ($p$) signals for alpha (red), beta (green) and gamma (blue) pairs**, scaled to $[-1, 1]$ range. The seed signals $s_i, i = 1, ...6$ were produced by passing white noise generated at 200 $s^{-1}$ sample rate through 8-th order Butterworth filter tuned to the corresponding frequency band. The evoked parts were constructed using expressions $c_{1,2} = \pm e(t)\sin(6\pi t/T)$ for the alpha pair and $c_{3,4} = -e(t)h(t, T)$ for the beta pair.

## 4. Discussion

The resulting set $\boldsymbol{p}$ shown in the third column in Fig. 1 has a correlation matrix exactly matching the target (8). However, one can notice that due to linear mixing of the seed signals in accordance with Eq. (5), some leakage of gamma frequencies into the alpha band pair $p_1,\ p_2$, and of alpha-frequencies into the gamma pair $p_5,\ p_6$ occurred. Such leakage cannot be avoided with this type of algorithm if one wants the signals from different bands to be partially correlated in the output. This should be kept in mind when using other, more complicated seed signals for modeling: one may expect specific distinct features of those to be mixed in the resulting set in accordance with the requested dependencies between signals described by $\hat{\boldsymbol{C}}_p$.

Also as explained in *section 2.2* a solution wouldn't exist if matrix (8) were not positively defined. In practice verifying this is very important because an arbitrarily constructed $\hat{\boldsymbol{C}}_p$ often turns out to be not positively defined and therefore cannot represent any realizable set of mutual correlations. For example in our case this happens if a stronger correlation between alpha and gamma pairs is requested – say 0.5 instead of 0.33. An intuitive explanation of this mathematical fact is that for high enough correlation between the alpha and gamma pairs the strong (anti) correlation among the alpha signals ($(\hat{\boldsymbol{C}}_p)_{1,2} = -0.9$) would induce a stronger correlation among the gamma signals, which contradicts the requirement that the gamma pair should be weakly correlated ($(\hat{\boldsymbol{C}}_p)_{5,6} = 0.2$).

## Appendix A. Mathematical proofs

Start with verifying that signals (3) satisfy constraints (4). First recall that $\langle \tilde{c}_i^2 \rangle = 1$ is not in fact a constraint but our original normalization assumption for the evoked signals, which can always be met. Next check that oscillatory parts have zero means: $\langle \boldsymbol{rE} \rangle = 0$. Indeed, according to (5) $\langle \boldsymbol{rE} \rangle = (\hat{\boldsymbol{C}}_p - \boldsymbol{Q}\hat{\boldsymbol{C}}_c\boldsymbol{Q})^{1/2}\boldsymbol{C}_{sPE}^{-1/2}\langle \boldsymbol{sPE} \rangle$, but $\langle \boldsymbol{sPE} \rangle = (1/m)\boldsymbol{sPE}\boldsymbol{1}_m^T = (1/m)\boldsymbol{sPe}^T = 0$, because by construction of the projector $\boldsymbol{P}$ in (6) $\boldsymbol{Pe}^T = 0$. Similarly oscillatory parts $\boldsymbol{rE}$ are uncorrelated with evoked signals $\boldsymbol{c}$: $\langle \boldsymbol{rEc}^T \rangle = \langle \boldsymbol{rE}\tilde{\boldsymbol{c}}^{\boldsymbol{T}} \rangle = 0$, because again by construction of $\boldsymbol{P}$ we have $\langle \boldsymbol{sPE}\tilde{\boldsymbol{c}}^T \rangle = 0$. Thus the first two conditions in (4) are verified. Additionally for the varying part of the evoked component of the total signal $\widetilde{\boldsymbol{Qc}}$ we have $\widetilde{\boldsymbol{Qc}} = \boldsymbol{Q}\tilde{\boldsymbol{c}}$, because $\boldsymbol{Q}$ does not depend on time. It follows then that the oscillatory part of $\boldsymbol{p}$ is uncorrelated with its evoked part: $\langle \boldsymbol{rE}\widetilde{(\boldsymbol{Qc})}^T \rangle = \langle \boldsymbol{rE}\tilde{\boldsymbol{c}}^T \rangle \boldsymbol{Q} = 0$.

The third constraint $\langle \tilde{p}_i^2 \rangle = 1$ will be satisfied automatically provided the target covariance $\hat{\boldsymbol{C}}_p$ is actually obtained, because $\hat{\boldsymbol{C}}_p$ is a correlation matrix. Let us check this now. By definition, covariance of $\boldsymbol{p}$ is $\langle \tilde{\boldsymbol{p}}\tilde{\boldsymbol{p}}^T \rangle$. We have

$$\langle \tilde{\boldsymbol{p}}\tilde{\boldsymbol{p}}^T \rangle = \langle (\boldsymbol{rE} + \boldsymbol{Q}\tilde{\boldsymbol{c}})(\boldsymbol{rE} + \boldsymbol{Q}\tilde{\boldsymbol{c}})^T \rangle = \langle \boldsymbol{rE}^2\boldsymbol{r}^T \rangle + \boldsymbol{Q}\langle \tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}^T \rangle \boldsymbol{Q}$$

because as we just showed cross-correlations between evoked and oscillatory parts disappear. Then using expression (5) for $\boldsymbol{r}$ and (7) for $\boldsymbol{C}_{sPE}$, we get

$$\begin{aligned}
\langle \tilde{\boldsymbol{p}}\tilde{\boldsymbol{p}}^T \rangle &= (\hat{\boldsymbol{C}}_p - \boldsymbol{Q}\hat{\boldsymbol{C}}_c\boldsymbol{Q})^{1/2}\boldsymbol{C}_{sPE}^{-1/2}\langle \boldsymbol{sPE}^2\boldsymbol{Ps}^T \rangle \boldsymbol{C}_{sPE}^{-1/2}(\hat{\boldsymbol{C}}_p - \boldsymbol{Q}\hat{\boldsymbol{C}}_c\boldsymbol{Q})^{1/2} \\
&\quad + \boldsymbol{Q}\hat{\boldsymbol{C}}_c\boldsymbol{Q} = \\
&(\hat{\boldsymbol{C}}_p - \boldsymbol{Q}\hat{\boldsymbol{C}}_c\boldsymbol{Q})^{1/2}\boldsymbol{C}_{sPE}^{-1/2}\boldsymbol{C}_{sPE}\boldsymbol{C}_{sPE}^{-1/2}(\hat{\boldsymbol{C}}_p - \boldsymbol{Q}\hat{\boldsymbol{C}}_c\boldsymbol{Q})^{1/2} + \boldsymbol{Q}\hat{\boldsymbol{C}}_c\boldsymbol{Q} = \\
&(\hat{\boldsymbol{C}}_p - \boldsymbol{Q}\hat{\boldsymbol{C}}_c\boldsymbol{Q}) + \boldsymbol{Q}\hat{\boldsymbol{C}}_c\boldsymbol{Q} = \hat{\boldsymbol{C}}_p
\end{aligned}$$

– end of proof.

## References

[1] X. Fan, H. Markram, A Brief History of Simulation Neuroscience, Frontiers in Neuroinformatics 13 (2019) 32. doi:10.3389/fninf.2019.00032.