Neurons in the mouse brain correlate with cryptocurrency price: a cautionary tale

Guido T. Meijer

Champalimaud Centre for the Unkown, Lisbon, Portugal
Correspondence: guido.meijer@research.fchampalimaud.org

20 August 2021

Abstract

In this paper I report the discovery of neurons which showed a neural correlate with ongoing fluctuations of Bitcoin and Ethereum prices at the time of the recording. I used the publicly available dataset of Neuropixel recordings by the Allen Institute to correlate the firing rate of single neurons with cryptocurrency price. Out of ∼40,000 recorded single neurons, ∼70% showed a significant correlation with Bitcoin or Ethereum prices. Even when using the conservative Bonferroni correction for multiple comparisons, ∼35% of neurons showed a significant correlation, which is well above the expected false positive rate of 5%. These results were due to 'nonsense correlations': when correlating two signals which both evolve slowly over time, the chances of finding a significant correlation between the two are much higher than when comparing signals which lack this property.

Introduction

In a typical neuroscience experiment we record a group of neurons, find neurons which show a neural correlate of a behavioral variable of interest, and restrict further analyses to this neuronal subset. The central assumption in this approach is that the neural correlate under scrutiny is not merely a statistical anomaly but a neural code that the brain actually uses. Often this is true, but there are cases in which this can lead to false conclusions. For example, when correlating neural activity with a behavioral variable which slowly evolves over time. Both neural activity and such a variable show temporal auto-correlations and are therefore very likely to result in a 'nonsense correlation' [1].

This statistical pitfall can result in the drawing of erroneous conclusions, as was argued to be the case when describing action-value coding neurons in the striatum [2]. It can also lead to amusing spurious findings: it was shown that a daily average of population activity in the rat motor cortex correlated with day-to-day fluctuations of stock prices. The temporal auto-correlation in the signals even allowed neural activity of these neurons to be used to predict the stock value of the next day [3].

Here, I aimed to illustrate this statistical pitfall at a large scale and investigated several of the proposed methods to circumvent the issue of nonsense correlations in the brain. I correlated spiking activity from tens of thousands of neurons in the mouse brain with ongoing fluctuations in the price of Bitcoin and Ethereum, the two most well-known cryptocurrencies. I found that ∼70% of neurons showed a significant correlation with cryptocurrency price at the time of the recording. This was not merely a multiple comparisons problem because when using the conservative Bonferonni correction, still a large fraction of neurons showed a significant correlation. Two methods, proposed by Kenneth Harris [1, 4], were successful in reducing the false alarm rate to acceptable levels. When analysing signals which slowly evolve over time, one should be aware of these statistical pitfalls and use the proper control analyses to correct for them.

Methods

I used the publicly available Visual Coding - Neuropixels dataset [5] provided by the Allen Institute as part of the Brain Observatory [6]. Briefly, spiking activity was recorded in posterior cortical and subcortical structures using high-density Neuropixel silicon probes. The dataset contained spiking activity of 40,010 neurons recorded in 58 mice which were head-fixed and passively viewing visual stimuli.

Single neuron activity was correlated with ongoing fluctuations in cryptocurrency price by binning each neuron's spike train in 60 second bins and calculating the spike rate in spikes per second per time bin. As a second method of defining neuronal activity, pseudo trials were generated by uniformly drawing 500 trial onset times from the entire length of the recording session. Subsequently, trials were defined as 300 ms windows after the onset times and spike counts were obtained for these time windows. The concurrent price of Bitcoin and Ethereum was collected using the Python library Historic-Crypto (`https://pypi.org/project/Historic-Crypto/`). Cryptocurrency prices were defined as the opening bid in 60 second bins starting at the exact start time and date of the recording. Spike rates and cryptocurrency prices were correlated using Pearson correlation.

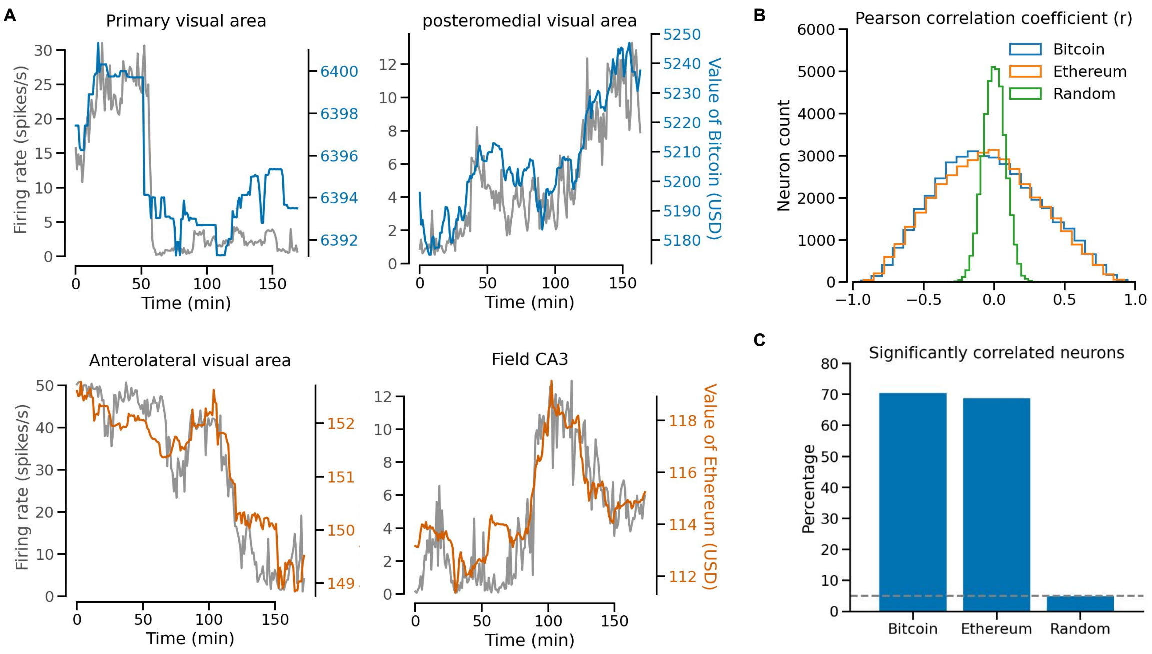

Figure 1: Neural correlate of cryptocurrency price. (**A**) Example neurons from four different brain regions which showed a strong correlation with ongoing price fluctuations of Bitcion (top) and Ethereum (bottom). Spike rate and cryptocurrency price was binned in 60 second bins. (**B**) Distribution of Pearson correlation coefficients was centered at zero but showed a large fraction of neurons that were positively and negatively correlated with cryptocurrency price. Correlation with a vector of random numbers was added as a control. (**C**) A high percentage of neurons showed a significant correlation with Bitcoin or Ethereum price ($p < 0.05$), in contrast only 4.9% of neurons were correlated with the random vector, close to the Type I false positive rate of 5% (grey dotted line).

The session permutation method was implemented by querying price vectors of equal duration from different points in time. To generate a null-distribution, I took 1000 price vectors starting at random time points between 2019 and 2020, and correlated these vectors with the spiking rate of each neuron using Pearson correlation. A p-value was defined as the fraction of times the Pearson correlation coefficient of the correlation between the neural activity and the price vector at the time of the recording was higher than the correlation with the price vectors from different times.

To perform the linear shift method, a time window of 50 minutes in the middle of the session was defined in which firing rate was correlated with cryptocurrency price using Pearson correlation. Subsequently, the window with the to-be-correlated metric is shifted throughout the session such that the neural activity is now correlated with the metric from a different part of the session. In this case, the null-distribution was generated by shifting the price window to earlier and later time points in the session in one minute steps while keeping the firing rate window the same. The p-value was defined as the fraction of times the correlation of the original price vector was stronger than the price vectors from shifted time windows.

Distinct temporal components were filtered out of the cryptocurrency price fluctuations using a 4th order Butterworth filter. Three different filters were used: a high pass filter which filtered out anything below 0.3 cycles per hour (cph), a band stop filter which took out frequencies between 0.3 and 2 cph and a low pass filter which removed anything above 2 cph. Each of these filtered price traces were subsequently correlated to binned spike trains of all neurons as described above.

Results

Many neurons showed a strong correlation in their firing rate with the price of Bitcoin or Ethereum at the time of the recording (Figure 1A). All plots of neurons with a correlation coefficient (r) of > 0.85 can be found here: https://figshare.com/articles/figure/Crypto-coding_neurons/14445480. The distribution of correlation coefficients between firing rate and cryptocurrency price over all neurons was very broad and included very strong correlations. To investigate whether any property of these price fluctuations made them particularly susceptible to spurious correlations, a vector with random numbers (a uniform random draw between 100 and 200) was correlated with the firing rate traces as a control; the random vector was only weakly correlated with firing rates (Figure 1B). For a remarkably large percentage of neurons, the correlation with cryptocurrency price was significant (Bitcoin: 70.5%, Ethereum: 68.8%) while the correlation with the random vector only resulted in 4.9% significantly correlated neurons which was around the expected Type I error rate of 5% (Figure 1C). These spurious correlations could not be controlled

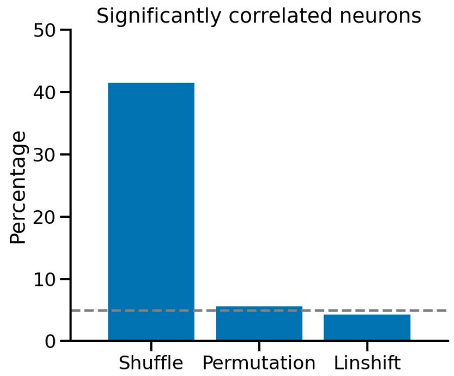

Figure 2: Methods to correct for nonsense correlations. *Shuffle*: shuffling the cryptocurrency price traces to determine significance did not reduce the false positive rate to the desired level of 5%. *Permutation*: constructing a null-distribution of correlations using 1000 price traces from other time points worked well to bring the false positive rate down to ∼5%. *Linshift*: using the *linear shift* method whereby a window of time points is shifted through the session to generate a null-distribution also reduced the amount of significantly correlated neurons to the desired level.

for by correcting for multiple testing. When using the conservative Bonferroni correction, a large fraction of neurons still showed a significant correlation with crypto price (Bitcoin: 34.7%, Ethereum: 33.9%).

Binning neuronal activity in 60 second bins generally does not happen in neuroscience. Therefore, I generated 500 pseudo trials with a duration of 300 ms, which is a commonly used time window, and obtained the spike counts of all neurons for these time windows. When correlating the spiking activity during these trials with the cryptocurrency price at that time, I still found a large fraction of significant neurons (Bitcoin: 60.1%, Ethereum: 59.2%; Bonferroni corrected: Bitcoin: 24.3%, Ethereum: 23.8%).

I investigated whether these spurious correlations could be controlled for using different methods to determine whether neurons were significantly correlated to cryptocurrency price. A commonly used method to determine significance is to shuffle one of the traces a large number of times and calculate the correlation coefficient for every shuffling iteration. This null-distribution of correlation coefficients is subsequently used to determine significance; the p-value is defined as the fraction of times the correlation of the original signal was stronger than the shuffled traces. This approach still resulted in 41.5% of significantly correlated neurons (Figure 2).

Next, I used two methods proposed by Kenneth Harris [1] which are specifically designed to control for temporal auto-correlations: the session permutation and the linear shift method. In the session permutation method one uses the data from other sessions, recorded under the same conditions, to generate a null-distribution from which a p-value can be derived. This resulted in 5.6% significant neurons which was close to the desired false positive rate of statistical testing.

The linear shift method [4] works by defining a time window in the middle of the session and correlating neural activity in this window with a metric of interest at that time, this window is then shifted throughout the session to generate a null-distribution. This method of controlling for nonsense correlations was slightly more conservative than the permutation method and resulted in 4.2% of significantly correlated neurons. Using these methods, a false discovery rate (FDR) correction of p-values sufficed to almost completely eliminate any false positives (permutation: 0.07%, linear shift: 0%).

What property of cryptocurrency price fluctuations was most important in eliciting these strong nonsense correlations? Cryptocurrency prices evolve over time at several different time scales (Figure 3A). Generally, there is a slow trend over time with fast price fluctuations on top. To investigate which of these components was the most important driver of nonsense correlations I filtered out the slow component (High pass), the medium component (Band stop) and the fast component (Low pass; Figure 2B) and correlated each of these filtered traces with the neural activity. Filtering out the slowly evolving trends resulted in a large drop of significantly correlated neurons while filtering out the medium and fast components resulted in a moderate increase of neurons that were correlated to these traces (Figure 3B). This suggests that the slow trends in cryptocurrency prices are the main driver of nonsense correlations.

Discussion

Why did such a large fraction of neurons show a significant correlation with crytocurrency price? We can rule out that neurons in the mouse brain actually encoded cryptocurrency price, as they did not have access to this information during the recording. Moreover, mice almost certainly lack the capacity to read and interpret complex financial data. The most likely explanation is that firing rates and cryptocurrency prices slowly evolve over time and the time constant of these temporal auto-correlations happened to be similar. Because the two signals that are being correlated to one another share this statistical property, the chances of finding a significant correlation between the two are much higher than the usual false positive rate of 5%. Specifically, the slow trends in cryptocurrency prices resembled neuronal activity patterns because taking out these slow trends strongly reduced nonsense correlations. Furthermore, the extremely large number of neurons in the dataset allowed for strong correlations to be observed purely by chance. Because of the temporal auto-correlations in both signals, conventional methods like multiple comparisons corrections failed to reduce the false positive rate to acceptable levels.

When analysing signals that slowly evolve over time, like neural activity, one should take the utmost care to avoid the pitfall of 'nonsense correlations' [1]. Although

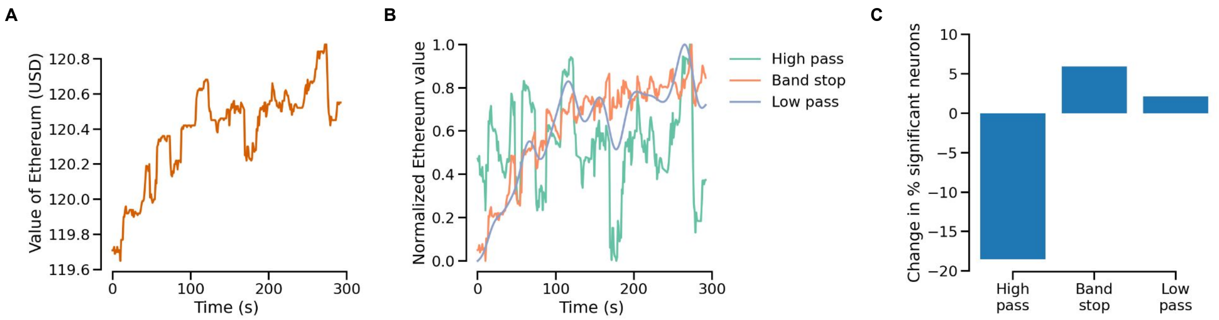

Figure 3: Influence of different temporal components on nonsense correlations. (**A**) Example cryptocurrency price of Ethereum over a time of five hours. (**B**) The price vector from **A**, filtered in three different ways: High pass (0.3 cycles per hour), Band stop (0.3 - 2 cph) and Low pass (2 cph). (**C**) Filtering out the slow component of the cryptocurrency price fluctuations resulted in a large decrease in the percentage of significantly correlated neurons. Band stop and Low pass filtered traces resulted in slightly more correlated neurons.

this issue is widely discussed in other scientific fields, its importance is only recently gaining traction in systems neuroscience. This paper serves as a cautionary tale that the potential confound of nonsense correlations is to be taken seriously. When not properly controlled for, it can lead to the misleading conclusion that 70% of neurons in the mouse brain encode cryptocurrency prices.

Data and code availability

The Visual Coding - Neuropixels dataset from the Allen Brain Institute: https://allensdk.readthedocs.io/en/latest/visual_coding_neuropixels.html.
The code used to analyse the dataset and generate the figures: https://github.com/guidomeijer/crypto-correlations

Acknowledgments

I thank the Allen Brain Institute for making their dataset publicly available. I thank Brandon Benson for the code for the linear shift method and Kenneth Harris who's papers on this topic inspired most of the thinking. I thank Alex Leighton for conversations and comments on the manuscript.

References

[1] K. D. Harris, Nonsense correlations in neuroscience, *bioRxiv* (2020).

[2] L. Elber-Dorozko and Y. Loewenstein, Striatal action-value neurons reconsidered, *eLife* **7**, e34248 (2018), ISSN 2050-084X.

[3] T. Marzullo, C. Miller, and D. Kipke, Stock Market Behavior Predicted by Rat Neurons, *Annals of Improbable Research* **12**, 401 (2016).

[4] K. D. Harris, A Shift Test for Independence in Generic Time Series, *arXiv:2012.06862* (2020).

[5] J. H. Siegle, X. Jia, S. Durand, S. Gale, C. Bennett, N. Graddis, G. Heller, T. K. Ramirez, H. Choi, J. A. Luviano, et al., Survey of spiking in the mouse visual system reveals functional hierarchy, *Nature* **592**, 86 (2021), ISSN 1476-4687.

[6] S. E. J. de Vries, J. A. Lecoq, M. A. Buice, P. A. Groblewski, G. K. Ocker, M. Oliver, D. Feng, N. Cain, P. Ledochowitsch, D. Millman, et al., A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex, *Nature Neuroscience* **23**, 138 (2020), ISSN 1546-1726.