Reviewer #1 (Kenneth Harris)

Great illustration of nonsense correlations, and glad the methods worked!

Also how often do you get a chance to read a carefully caveatted statement like "mice almost certainly lack the capacity to read and interpret complex financial data" in a scientific paper.

Thank you for your kind words Kenneth!

Reviewer #2

The preprint aims to warn neuroscientists from the use of standard correlation methods, such as the Pearson correlation coefficient, when comparing slowly changing time series such as cryptocurrency prices against binned (with 60 s long time bins, almost never used in neuroscience studies) neuronal spiking activity, to estimate neuron tuning. Application of this method leads to a high number of false positive (FP) tuned neurons, even when standard multiple comparison correction methods are applied. In contrary, application of more sophisticated techniques like the session permutation and linear shift method (both extensively described in Harris 2020, biorXiv), leads to a reasonable estimate of FP tuned neurons.

I think the notion that the Pearson correlation coefficient should be taken merely as a first pass indication of potential relation between neuronal firing and variable of interest, requiring validation by ad hoc methods like cross-validated statistical models and linear shift, has now been around the field for many years (one out of many examples: Campagner et al. 2016, elife) and recently reviewed in Harris 2020,biorxiv (as mentioned by the author).Therefore, despite being an interesting exercise, I don't see the relevance nor the novelty of this work.

The reviewer raises a relevant point: 60 second time bins are never used in neuroscience studies. Therefore, I performed an additional analysis where I generated 500 pseudo trials which were randomly interspersed throughout the session. These trials were time windows of 300 ms long, which is a commonly used trial length in neuroscience. When correlating the spike counts during these trials with the concurrent price of Bitcoin I found that 60.1% of neurons showed a significant correlation (24.3% after Bonferroni correction) which is still much higher than the expected false positive rate of 5%. A paragraph was added to the results section of the revised manuscript about these results.

Another point which is raised by the reviewer is that a correlation is merely a first pass indication of a relationship which requires ad hoc validation. I agree with this but I would point out that even cross-validated statistical models can be influenced by temporal autocorrelations, depending on how the experimenter does the cross-validation. Imagine that there is strong temporal auto-correlation and the experimenter uses a leave-one-out cross-validation. In this case the experimenter will probably find above chance performance of their model because the trial before and after the trial taken out by the cross-validation are still in the training set and these trials are not independent from the to-be-evaluated trial because of the auto-correlations.

The last main point raised by the reviewer regards the relevance of this work. Although I agree that the points made in this paper are not novel (no claims of novelty are made in the manuscript), I would argue that they are relevant. In my experience, there is insufficient awareness of the issue of temporal auto-correlations in the field of systems neuroscience. While in other related fields, such as the fMRI field, this issue is broadly acknowledged and controlled for. There are indeed many methods available to control for temporal auto-correlations, however, I find the apprehension of this statistical pitfall lacking and hope to increase awareness of this issue.

Reviewer #3 (Anirudh Kulkarni)

The study by Meijer presents an example of the erroneous conclusions that can be reached by using correlations without any systematic control analyses. More specifically, one could be led to the false conclusion that the neural activity in the mouse brain correlates with the fluctuations in the Bitcoin and Ethereum price while not using the appropriate control analyses. The study demonstrates the usefulness of two methods developed recently by K. D. Harris(2020) namely the Permutation method and the LinShift method to test the significance of correlations rather than the standard ‘shuffle’ method that is used in literature. This article serves as a good warning when correlating neuronal activity with any variables including behavioural data.

The article is well written. The abstract clearly represents the methodology used and aims of the article. The introduction highlights the problem in current neuroscience data analyses methods by showing how 'nonsense correlations' in data can lead to false conclusions. It sets out to use the proposed methods in literature to see if they were successful in reducing these false errors that creep into the analysis. The article uses the publicly available spiking activity of 40,010 neurons in 58 mice available from the Allen Institute and the Bitcoin and Ethereum prices using the Python library Historic-Crypto. The Pearson correlation coefficient between the single neuron activity and the cryptocurrency prices have been calculated. The methodology is thus sound and replicable.

Thank you for your positive feedback on this manuscript.

Corrections: For Figure 1, it would be good to use consistent colours for the different entities over all the panels: i.e. blue represented the firing rates of the neurons in Fig. 1A whereas it represents Bitcoins in Fig. 1B.

This is a good suggestion, I changed the colors in the revised manuscript so that they match over panels.

The methods described in the results section i.e. the permutation method and the linear shift method could ideally be moved to the methods section to allow a smoother reading of the results.

Done, thank you for this suggestion.

The statistical test used to compare the different distributions in Figure 2 and to calculate the p-values needs to be stated explicitly in the methods section.

I am not sure I fully understand this point, no statistical test was used to derive the p-values that are reported in Figure 2. In case you meant to explicitly state that Pearson correlation was used, I added this to the revised manuscript. The p-values were calculated using the following approach: a correlation was done between the neural activity and the actual crypto price which resulted in a certain correlation coefficient (r). Subsequently, the relationship between neural activity and the crypto price was destroyed by either taking the price from a different point in time (session permutation) or taking the price from a different point in the same session (linear shift), this was done many times resulting in a null-distribution of correlation coefficients. A p-value was determined by calculating the fraction of r-values in the null-distribution that were higher than the original r value. The logic behind this is that if the actual correlation is significant, it would fall in the upper 95th percentile of the null-distribution and therefore there will be <5% of values in the null-distribution are higher than the original value.

The discussion is well guided and comments on the results obtained. The underlying cause of the results are, however, not explicitly clarified and could be further looked into by testing the hypothesis whether using signals with temporal autocorrelations that have different time constants still maintains significant correlations.

This is a good point. I investigated which temporal components of the cryptocurrency price traces were most important for driving nonsense correlations by filtering out different frequencies from the price traces and looking at how this influenced the amount of correlated neurons (Figure 3 in the revised manuscript). I found that the slow component of the price fluctuations was the main driver of nonsense correlations because selectively removing the slow drift resulted in a sharp drop of significantly correlated neurons. Interestingly, removing the medium and fast frequencies actually resulted in even more significantly correlated neurons.

Minor corrections:

Abstract: line 4 ~40.000 should be ~40,000

Methods: line 6 40.010 must be 40,010.

Thank you, this was amended.