Peer Review File

Manuscript Title: General-relativistic precession in a black-hole binary

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

A. Summary of the key results

The authors reanalyze the gravitational wave data recorded by the LIGO and Virgo detectors from binary black hole merger GW200129_065458, hereafter GW200129. The motivation of this reanalysis is that the LIGO and Virgo collaborations estimated the parameters of GW200129 using two GW template banks that are suitable for the full binary black hole parameter space, but are not the best available at the high mass space where GW200129 is located. In this mass space, assuming some other applicable constraints on the binary parameters, the template bank adopted by the authors, NRSur7dq4, is a more accurate description of the expected GW signal. With this reanalysis, the authors find strong support (Bayes factor 30:1) of non-zero black hole spin component perpendicular to the binary's orbital axis (hereafter precessing spin), which results in orbital precession. Previously, there has been strong evidence found for the presence of precessing spin in the population of black hole mergers detected by LIGO/Virgo, and some evidence for several black hole mergers having precessing spin. Nevertheless, the manuscript presents the first high significance identification of precessing spin for a single black hole merger.

B. Originality and significance: if not novel, please include reference

The manuscript carries out a standard parameter estimation analysis regularly performed by LIGO/Virgo and others for detected events, often with multiple template banks and with other variations in the analyses, such as prior assumptions. The presented analysis represents two main changes compared to the standard LIGO/Virgo analysis for GW200129. First, it uses the template bank NRSur7dq4 instead of the two other ones used by LIGO/Virgo for this event. NRSur7dq4 was used by LIGO/Virgo and others for a number of binary mergers, but not specifically for GW200129. Second, the authors carry out a detailed investigation of possible sources of error, which is more detailed than typical for typical, although not all LIGO/Virgo analyses. Parts of this analysis are novel as it aims to demonstrate the accuracy of the results and why NRSur7dq4 is better and more accurate than the template banks adopted by LIGO

Assuming the high significance holds (see below), the demonstration of precessing spin at high confidence for an individual black hole merger is a new, nice demonstration of GR-related precession and is evidence for the unusual formation mechanism of GW200129. The astrophysical significance of the finding, however, is limited by the fact that the presence of precessing spin has already been demonstrated at high confidence for the detected mergers as a population. This population-wide result is key in understanding how the binary mergers LIGO/Virgo is detecting are formed. In addition, some of LIGO/Virgo's binaries are interesting due to their special properties, for example high mass. This includes among others GW190521. For this event, the demonstration of precessing spin adds to the peculiarity of the event, and is a further hint at its formation, even though the presence of precessing spin is not as high confidence as demonstrated in the manuscript here. However, since GW200129 is not "special" in other ways, it is not clear how the new information presented in the manuscript furthers our astrophysical understanding of black hole mergers. The authors essentially say this when they list the possible formation channels for GW200129, while stating that nothing conclusive can be said from this event.

Data & methodology: validity of approach, quality of data, quality of presentation

The authors present a nice and thorough analysis of the data where they take particular care to identify possible sources of error that can lead to biased parameter estimation. The main parameter estimation they carry out uses standard LIGO/Virgo software and settings.

The authors adequately establish that noise artifacts and template uncertainties do not affect the results meaningfully. I see, however, two remaining issues with the claimed significance of the results.

1. The first one is related to the assumed prior assumptions on the binary parameters. The authors do not explicitly state their priors (they should), but from the description I assume this is the same what the official LIGO/Virgo analysis used. For some other events, such as GW190521, the role of the priors has been investigated in a number of studies, and these studies found large variations in the estimated parameters based on the prior assumptions. It is not inconceivable that a similar dependence is present here. This could be addressed by considering other prior distributions and demonstrating that they still result in a large Bayes factor favoring precession.

2. Orbital eccentricity can also alter the obtained results, potentially significantly. In fact it was found that eccentricity can mimic precessing spin when parameter estimation is carried out with a non-eccentric template bank. The authors briefly mention this, and given the limited templates available for eccentric waveforms it is difficult to test this assumption. Nevertheless, without taking eccentricity into account, it is not clear that a high confidence claim is justified. Assuming that eccentricity can indeed mimic precessing spin, the question then becomes whether at least 1 in 30 black hole mergers are eccentric. This is consistent with existing model predictions.

C. Appropriate use of statistics and treatment of uncertainties

In general the statistical presentation appears accurate, and most uncertainties are taken into account, see the previous point for two exceptions.

As an additional point, it was not entirely clear how the confidence regions for the binary mass were determined. In line 175 the authors find 99% support above 58 solar mass, but this result was obtained by limiting the mass prior to above o 56.7. It is not clear whether this initial limit could change the result. For instance if some more unequal mass binary was also consistent with the results, the presented analysis could not find this. For example, for GW190521 multiple studies found that a more unequal mass ratio is also consistent with the data, and found that for certain prior assumptions it is even favored.

E. Conclusions: robustness, validity, reliability

Overall, there are two main and one minor possible issues with the robustness of the results. Other than these, the rest of the analysis is robust.

F. Suggested improvements: experiments, data for possible revision

I suggest the authors investigate whether prior assumptions could alter their result.

G. References: appropriate credit to previous work?

Appropriate references are provided in the manuscript.

H. Clarity and context: lucidity of abstract/summary, appropriateness of abstract, introduction and conclusions.

In general, the presentation of the results is of high quality and mostly understandable with small caveats: there is no information in the main text on what improved compared to the standard LIGO/Virgo results. It would also be useful for the readers to see the parameter estimation results for the two LIGO/Virgo templates in comparison to those found in the manuscript.

Author Rebuttals to Initial Comments:

We thank the referee for their detailed reading of our paper. We address the issues raised in the refereeâ€™s report, and argue that Nature is the appropriate journal for a result of this importance.

B. Originality and significance: if not novel, please include reference

Assuming the high significance holds (see below), the demonstration of precessing spin at high confidence for an individual black hole merger is a new, nice demonstration of GR-related precession and is evidence for the unusual formation mechanism of GW200129. The astrophysical significance of the finding, however, is limited by the fact that the presence of precessing spin has already been demonstrated at high confidence for the detected mergers as a population. This population-wide result is key in understanding how the binary mergers LIGO/Virgo is detecting are formed. In addition, some of LIGO/Virgoâ€™s binaries are interesting due to their special properties, for example high mass. This includes among others GW190521. For this event, the demonstration of precessing spin adds to the peculiarity of the event, and is a further hint at its formation, even though the presence of precessing spin is not as high confidence as demonstrated in the manuscript here. However, since GW200129 is not â€œspecialâ€ in other ways, it is not clear how the new information presented in the manuscript furthers our astrophysical understanding of black hole mergers. The authors essentially say this when they list the possible formation channels for GW200129, while stating that nothing conclusive can be said from this event.

RESPONSE: The scientific importance of our analysis of GW200129 rests on this being the first observation of an important physical phenomena, namely strong-field general-relativistic orbital precession. It is *not* based on the confirmation that binaries can form with mis-aligned spins. As the referee notes, compelling evidence for this conclusion has already been published in previous analyses conducted by the LIGO-Virgo-Kagra (LVK) collaborations, and indeed earlier in work by ourselves and others [Farr, W., et al., Nature 548, 426-429 (2017), Tiwari, V., et al., ApJ 868, 140 (2018), Hoy, C. et al., ApJ 928, 75 (2022)]. We have modified the abstract to remove any implication that the main focus of our results might be the properties of the binary population.

We wish our result to be published in Nature because spin induced orbital precession has *not* been unambiguously measured in a GW signal prior to GW200129. This makes our analysis a landmark result in GW astronomy. Indeed, the first measurement of precession in a GW signal has been a major long-standing scientific prize. Among many illustrations of this point, we note that a measurement of precession in a GW signal has been deemed an exceptional discovery by the LVK (see page 15 of the LIGO scientific program [1]), putting our analysis on the same footing as other exceptional discoveries such as GW150914, GW170817, GW190521 and GW190814. We also note that in a recent review talk on GW astronomy given in December 2022 [2], a measurement of precession in a GW signal was one of the five key outstanding unobserved phenomena from the latest LVK observing run (this talk also referred to an independent re-analysis of the LVK data, which also failed to identify the precession effects in this signal, a further illustration that our result is far from a routine data-analysis exercise).

To further illustrate the importance and significance of this result, we calculated that a GW200129-like signal (a binary with a spin magnitude greater than 0.8 and a spin mis-alignment greater than 70 degrees with measurable precession in the observed GW signal) will likely be observed only once in every 1200 observations. As such, we do not expect to observe an event like GW200129 in the next two observing runs (we have added text to the manuscript explaining this result, see L160-166). This is despite O4 and O5 likely including over 1000 observations. This expectation is based on the inferred population from Ref. [10] of our paper. Alternatively, it may be more likely that GW200129 represents a new class of signal which will present a significant challenge to the standard formation mechanisms. In the majority of observed BBH signals the SNR has not been high enough for precession to be measured, even if the black-hole spins are large and mis-aligned, and so these may be more common than previously thought.

This work is a natural successor to two other important related results that appeared in Nature: the first measurement of weak-field precession in 2003 [Burgay, M., et al., Nature 426, 531â€"533 (2003)], and the first indirect evidence for mis-aligned spins in 2017 [Farr, W., et al., Nature 548, 426-429 (2017)]. We now present the first direct evidence for the striking physical effect that general relativity predicts follows from spin misalignment, which is ten orders of magnitude faster than in the binary-pulsar measurements.

Data & methodology: validity of approach, quality of data, quality of presentation

1. The first one is related to the assumed prior assumptions on the binary parameters. The authors do not explicitly state their priors (they should), but from the description I assume this is the same what the official LIGO/Virgo analysis used. For some other events, such as GW190521, the role of the priors has been investigated in a number of studies, and these studies found large variations in the estimated parameters based on the prior assumptions. It is not inconceivable that a similar dependence is present here. This could be addressed by considering other prior distributions and demonstrating that they still result in a large Bayes factor favoring precession.

RESPONSE: We used priors that were in line with Appendix B.1 in [Abbott, B.P., et al., Phys.Rev.X 9, 031040 (2019)], i.e. flat in spin magnitude and the cosine of the misalignment angle, since these priors are the most agnostic choice (we thank the referee for pointing out that we did not explicitly state our priors and text has been added to the manuscript, see L193-200 and L367-368). Although there is no clear motivation to use prior distributions other than the agnostic choice because a) there are not enough observations to make a clear statement about the population of black holes in the universe and b) the formation mechanism of binary black holes is currently not known, we performed model selection with spin priors based on the most likely population inferred from the latest (O3) LVK analysis [Abbott, R., et al., arXiv:2111.03634]. This population-based prior strongly prefers low-spin binaries, and therefore disfavours the high spin values that we recover with standard priors. Nevertheless, when analysing GW200129 with this population based prior, we found that the inference again prefers large in-plane spins but the LVK population based prior is disfavoured by a factor of 5:1. We thank the referee for highlighting this and we have included text in the manuscript which describes this finding (see L362-383).

Although we only performed model selection with spin priors based on the most likely population inferred from the latest (O3) LVK analysis, we note that alternative choices for the spin priors could have been used, including e.g. a prior that strongly prefers high-spin binaries. For such choices we would expect the evidence for precession to increase significantly compared to the standard priors since it is evident that the data prefers larger black hole spins lying in the plane of the binary. Although we are free to choose a prior that can greatly enhance the odds of precession and will likely be preferred in the data, we consider the agnostic priors to provide the most well-motivated estimate of the significance of the precession measurement. We consider an in-depth analysis of a wider set of priors, as was done for e.g. GW190412 in follow-up studies after the main first analysis [Zevin, M., et al., ApJL 899, L17 (2020)] to be beyond the scope of this paper.

2. Orbital eccentricity can also alter the obtained results, potentially significantly. In fact it was found that eccentricity can mimic precessing spin when parameter estimation is carried out with a non-eccentric template bank. The authors briefly mention this, and given the limited templates available for eccentric waveforms it is difficult to test this assumption. Nevertheless, without taking eccentricity into account, it is not clear that a high confidence claim is justified. Assuming that eccentricity can indeed mimic precessing spin, the question then becomes whether at least 1 in 30 black hole mergers are eccentric. This is consistent with existing model predictions.

RESPONSE: As the referee notes, eccentricity and precession effects can in principle be confused in a GW observation. This is because both lead to modulations in the GW amplitude and phase. However, this confusion will most often occur when the observed GW is short such that only the merger is observed. This is because at merger both the eccentricity and precession modulations can occur on the orbital timescale. However, if the signal is long such that there is sufficient information in the inspiral, precession and eccentricity cannot be confused with one another since the precession modulations in the inspiral occur at a much longer timescale compared to the eccentricity modulations, which remain on the orbital timescale. This is one of the main reasons why it is possible that precession and eccentricity could be confused in GW190521 [Romero-Shaw, I. et al, ApJL 903, L5 (2020)], because that system is at roughly twice the total mass of GW200129, with no more than one orbit before merger measurable in the data. In contrast, GW200129 has nine orbits (or 18 GW cycles) present in the GW data. This means that during the late inspiral there would be nine eccentricity modulations, but only one precession modulation. This means that eccentricity and precession effects will *not* be mistaken in GW200129. We thank the referee for highlighting this and we have updated the text in the manuscript to include this argument, see L311-321.

C. Appropriate use of statistics and treatment of uncertainties

As an additional point, it was not entirely clear how the confidence regions for the binary mass were determined. In line 175 the authors find 99% support above 58 solar mass, but this result was obtained by limiting the mass prior to above o 56.7. It is not clear whether this initial limit could change the result. For instance if some more unequal mass binary was also consistent with the results, the presented analysis could not find this. For example, for GW190521 multiple studies found that a more unequal mass ratio is also consistent with the data, and found that for certain prior assumptions it is even favored.

RESPONSE: We understand that prior choices can prevent us from exploring the full parameter space, which could cause potential multi-modal structures in the posterior to be missed. As the referee notes, this was apparent in GW190521 with e.g. [Nitz, A. and Capano, C. ApJL 907, L9 (2021), Estelles, H., et al., ApJ 924, 79 (2022)] showing that this event is also consistent with an unequal mass ratio binary and that prior choices imposed on the mass ratio in the original LVK analysis did not allow for this region to be probed. We do not believe this to be an issue in our analysis for three reasons: first, GW200129 is a high mass binary observed with the highest SNR of any BBH to date with SNR ~27. For high-mass binaries, the total mass is the best-measured quantity (chirp mass is best measured for low-mass binaries). If there was a bimodality in the total mass, it is therefore likely to be an issue with the sampler, waveform systematics or noise artifacts. As detailed in the paper, we have performed extensive tests that our results are robust against all of these issues. Secondly, a recent paper that also analyzed GW200129 with the NRSur7dq4 waveform model [Varma, V., et al., arXiv:2201.01302] but without constraints on the total mass found consistent results to oursâ€™. Thirdly, the original LVK analysis analyzed GW200129 with no constraints on the total mass (and mass ratio extending to q = m1/m2 = 20, see Appendix E.3 of [Abbott R., et al., arXiv:2111.03606]) and this analysis also found no hints of additional structure outside of the prior bounds that we imposed in our analysis.

H. Clarity and context: lucidity of abstract/summary, appropriateness of abstract, introduction and conclusions.

In general, the presentation of the results is of high quality and mostly understandable with small caveats: there is no information in the main text on what improved compared to the standard LIGO/Virgo results. It would also be useful for the readers to see the parameter estimation results for the two LIGO/Virgo templates in comparison to those found in the manuscript.

RESPONSE: Although our Bayesian analysis has similarities to what was done for the LVK results, the significant improvement is in the interpretation and the additional analyses discussing the reliability of our results; both of which were missing in the LVK paper and indeed in other analyses undertaken by groups outside the LVK. As documented in the main text of our manuscript (L68-72), unlike the LVK results, we undertook numerous additional analyses to discuss the effects of noise and parameter estimation uncertainties on our interpretation of GW200129. As discussed at length in the Methods section, we also studied waveform model uncertainties, which were not mentioned in the LVK results. We thank the referee for highlighting that this is not mentioned in the main text of the paper and we have added a sentence to the paragraph starting on L63 mentioning this.

The referee also notes that some readers may find it useful to compare our parameter estimation results with those from the two LVK templates. We chose to not include them in our manuscript since, given the large systematic uncertainties reported by the LVK, we did not feel that it was useful to the reader to present them in detail in the same table as our parameter estimates.

References:
[1] LIGO Scientific Program, https://dcc.ligo.org/public/0168/M2000130/004/LSCprogram2020-21.pdf, p. 15.
[2] A. Nitz, "Review of Recent CBC search results", https://gwpaw2021.aei.mpg.de/program-item/recent-results-from-gravitational-wave-data-soc-members-brown-allen/.

Reviewer Reports on the First Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

I would like to thank the authors for the detailed responses to the raised concerns, and the improvement of the manuscript on several fronts. I am afraid, however, that my two main concerns have not been not alleviated:

1. Originality and significance: As there has previously been weak evidence for individual precession and strong evidence for population precession, the indication of precession for an individual event is not striking. In addition other than the statement that there is precession this does not substantially add to our astrophysical/physical understanding of the universe.

2. The authors additional analysis of the spin prior effects is interesting (the main point that LVK spins as prior do not support the main conclusion appear in the response but not in the updated manuscript). It also confirms that the conclusion of the manuscript depends on the assumed spin priors, weakening the main result. While the 5:1 support for flat spins is also interesting, I would caution that the analyzed event was selected specifically because of its likely high spin for the analysis, therefore I do not consider this information indicative of anything. If we pick 90 random samples from a low spin distribution, some of these will likely give similar low probabilities for low spin, despite being from that distribution.

Referee #3 (Remarks to the Author):

Manuscript 2022-01-00364A-Z 1st Report

This manuscript analyzes LVK GW signal 200129 using the NRSur7dq4 model whose greater accuracy for high-mass, low mass-ratio systems is required to properly characterize high SNR events. It demonstrates that this event favors the precessing hypothesis by a factor of at least 30:1, far greater than any previous LVK GW signal. This detection of precession was possible because of three features of this event: (1) the high network SNR, the low total mass allowing a full precession cycle within band, and (3) the small mass ratio (q ~ 0.6) allowing for a greater misalignment between the orbital and total angular momenta than for near equal-mass systems. Although the entire population of BBH mergers cannot be tightly constrained by this one event, the fact that precession has been observed in one of the few events in GWTC-3 with the features listed above allowing for such an observation suggests that current models predicting such precession in only 1 out of every 1200 events may need to be significantly revised. The quality of the text and figures was excellent. The uniqueness of this event and as a golden example of a precessing system, as well as its potential astrophysical implications, fully justifies publication in Nature (in my opinion). I had several minor suggestions listed below which the authors should consider, particularly point 5 which I believe is a mistake.

Suggestions:

1. It might be useful to describe the constraints on the secondary spin in more detail than “undetermined” in Table 1. GWTC-3 lists constraints on both chi_eff and chi_f, both of which have some dependence on the secondary spin.

2. At lowest PN order in the single-spin case, the angle between the BH spin and orbital angular momentum is preserved during the inspiral, and the orbital angular momentum tilts away from the total angular momentum. This is inconsistent with the claim below Fig. 1 (lines 109 – 111) that “The binary’s orbital plane is inclined to the total angular momentum by ~0.5 rad; this inclination does not change significantly through the late inspiral.” Could the authors clarify whether the evolution of the angle between the orbital and total angular momentum is also consistent with this expected single-spin behavior (within the given sensitivity).

3. The text states at the top of p. 8 (lines 201 – 202) that “our prior was flat in spin magnitude and the cosine of the misalignment angle.” I presume that it was also flat in the azimuthal angles of the two spins, which certainly affect the precession amplitude if the secondary BH is spinning. This should be stated for comprehensiveness.

4. Please clarify the sentence on p. 8 (lines 235 – 237) that “The signal-based precession frequency as measured from the theoretical model also contains oscillations that we know are not present in the orbital dynamics, and we filter those out of the results shown in Fig. 1.” Are these oscillations a predict of GR or an artifact of the theoretical model breaking down? At what confidence is it known that these oscillations are not present in the observed orbital dynamics?

5. Is the inequality in Eq. (1) backwards? Won’t two waveforms be distinguishable if their mismatch is above a certain threshold? Shouldn’t using only four degrees of freedom require a less stringent criterion for distinguishability (lower minimum mismatch for distinguishability).

Author Rebuttals to First Revision:

We first respond to the comments from the two available referees, and then to the comments and suggestions from the editor.

Response to Referee #1.

We illustrated in our previous response, with reference to the LIGO collaboration's scientific programme and other public documents, that the first clear measurement of precession in an individual GW signal would be seen as a major milestone in GW astronomy. Referee #1 did not address this point.

We have rewritten the Summary Paragraph to make the scientific value and novelty of this result clearer within the context of the field. The first three sentences now read, "The general-relativistic phenomenon of orbital precession has not yet been observed in strong-field gravity. Gravitational-wave observations of black-hole binaries are prime candidates, since we expect the astrophysical binary population to contain precessing binaries [1,2]. Imprints of precession have been investigated in several signals [3-5], but no definitive identification of orbital precession has been reported in any one of the 84 BBH observations to date [5-7] by the Advanced LIGO and Virgo detectors [8,9]."

In response to the referee's two points:

1. The evidence for precession in previous individual events has been either extremely weak or overshadowed by systematics; we mention them only because they are cases where the potential imprint of precession was searched for, but not conclusively found. And as noted previously, yes, the currently favoured black-hole-binary population models tell us that we can expect precessing binaries in the population, but that is quite distinct from the first observation of one, particularly when preceded by 83 non-observations. We have clarified in the Summary Paragraph the state of the field, and what our work shows.

2. We would like to clarify that in our previous response we did not state that results using the LVK spins as priors do not support the main conclusions of the paper; on the contrary, although the LVK spin distribution has very little support for high spins, analysis with those priors nonetheless recovers high spin and precession. To summarise, we do not consider the priors discussion to weaken the main result in any way. We have argued that we have used the most well-motivated and most agnostic priors for our study, and indeed the priors used in all LIGO-Virgo-Kagra (LVK) observations. We have now made this point explicit in the first paragraph of the main text.

We agree with the referee that the example of a 5:1 preference for a flat spin-magnitude prior over the LVK spin distribution, while interesting, does not have any bearing on the results here, and we have removed that discussion.

Response to Referee #3.

We thank Referee #3 for their positive comments, and are pleased that they share our excitement over this result.

To address their other points:

1. Secondary spin. We have added further explanation for our statement that the secondary spin is underdetermined.
Both the spin magnitude and tilt angle posteriors are essentially flat, and therefore the 90% credible intervals can
be any interval that covers 90% of the possible range of each of these parameters. We have provided the results of
this calculation in the table caption, to illustrate that this is indeed what happens.

2. As the referee notes, the angle between the total angular momentum and the orbital plane does increase over the course of the inspiral; the orbital angular momentum decreases but the in-plane spin magnitude remains constant, so the angle between them grows. But this change is very small over the short duration of the signal measurable in the detector. We agree that the statement as written was misleading, and we have changed it to, "although this angle increases through inspiral~\cite{Apostolatos:1994mx}, it does not change significantly over the short duration of this observation".

3. We have added a statement to clarify that the priors for the azimuthal angles are flat.

4. The oscillations in our calculation of these angles are due to artifacts in the model. These appear when we infer the precession dynamics from the waveform multipoles, but are not large enough to impact the accuracy of the model when measuring the properties of the signal. We know that these are unphysical because they are present in configurations where we know, from numerical-relativity simulations, that oscillations are *not* present in the dynamics. We have added a clarification that we know that the oscillations are not physical via comparisons with numerical-relativity results.

5. Yes, this should be the other way around! We have fixed this.

Other modifications, following comments and suggestions from the editor.

Following the editor's recommendation, we have removed the first two paragraphs of the main text. This indeed gets us to the main points of the paper much more quickly, and what previously seemed like essential introductory paragraphs were clearly unnecessary!

We have removed instances of "strong evidence" in the Summary Paragraph and in the main text. We used "strong evidence" in the sense of standard classifications of Bayesian evidence, but agree that this is unclear without additional context, which is not needed at this point in the paper.

We have removed several references from the main part of the paper. There are now 34. We feel that removing more would risk failing to properly cite essential work, but we can attempt to remove additional references if necessary.

We have given the table and figures titles, and relabelled the figures in the Methods section as "Extended Figures".

We have now provided statements about individual author contributions.

We have added links that will be used to allow readers to access our data. The link in the paper will be publicly available when the paper is published. For now the documentation are at https://hoyc1.github.io/GW200129-precession/, and the data files are at https://zenodo.org/record/6672460#.YrDRxbaZOlM.

Reviewer Reports on the Second Revision:

Referees' comments:

Referee #3 (Remarks to the Author):

I have reviewed the changes made to the manuscript in response to the previous referee reports. These changes have addressed most of the concerns expressed in my previous report. However, to fully respond to point 5 in my previous report, I believe that line 228 of the latest version should be changed from “stronger” to “weaker”. The condition M > 3.89/rho^2 is weaker than M > 6.68/rho^2, as not all binaries satisfying the former will satisfy the latter. A weaker condition is expect for a model with fewer free parameters (four vs. eight). Line 230 should similarly be changed from “more” to “less” stringent criteria. Also, to be grammatically correct, the plural “criteria” should be changed to the singular “criterion” on both lines 228 and 230 since there is a single inequality for the mismatch in both models.

After these small additional corrections are completed, the manuscript should be suitable for publication in Nature in my opinion.

Author Rebuttals to Second Revision:

Referee 3 raised one point on the text following Eq. (1). We agree that the presentation of the distinguishability inequality made the “weaker”/”stronger” wording confusing. We have now swapped the inequality sign so that rather than Eq. (1) representing a condition for two signals to be distinguishable, it is now a condition for them to be indistinguishable. The text has been modified to reflect that change, and whenever we refer to an inequality, it is always a “less than” inequality to define indistinguishability in line with what is now in Eq. (1). It now makes sense to say that a tighter constraint on the mismatch is a “stronger mismatch criteria”.

This analysis of model accuracy is crucial to our results and we thank the referee for helping us to make the presentation clearer.