



HAL
open science

Strong constraints on the gravitational law from *Gaia* DR3 wide binaries

Indranil Banik, Charalambos Pittordis, Will Sutherland, Benoit Famaey,
Rodrigo Ibata, Steffen Mieske, Hongsheng Zhao

► **To cite this version:**

Indranil Banik, Charalambos Pittordis, Will Sutherland, Benoit Famaey, Rodrigo Ibata, et al.. Strong constraints on the gravitational law from *Gaia* DR3 wide binaries. *Mon.Not.Roy.Astron.Soc.*, 2024, 527 (3), pp.4573-4615. 10.1093/mnras/stad3393 . hal-04277267

HAL Id: hal-04277267

<https://hal.science/hal-04277267v1>

Submitted on 30 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Strong constraints on the gravitational law from *Gaia* DR3 wide binaries

Indranil Banik¹,^{*} Charalambos Pittordis,² Will Sutherland,² Benoit Famaey³, Rodrigo Ibata,³ Steffen Mieske⁴ and Hongsheng Zhao¹

¹Scottish Universities Physics Alliance, University of Saint Andrews, North Haugh, Saint Andrews, Fife KY16 9SS, UK

²The School of Physical and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK

³Université de Strasbourg, CNRS UMR 7550, Observatoire astronomique de Strasbourg, 11 rue de l'Université, 67000 Strasbourg, France

⁴European Southern Observatory, Alonso de Cordova 3107, Vitacura, Santiago, Chile

Accepted 2023 November 1. Received 2023 October 28; in original form 2023 March 14

ABSTRACT

We test Milgromian dynamics (MOND) using wide binary stars (WBs) with separations of 2–30 kAU. Locally, the WB orbital velocity in MOND should exceed the Newtonian prediction by ≈ 20 per cent at asymptotically large separations given the Galactic external field effect (EFE). We investigate this with a detailed statistical analysis of *Gaia* DR3 data on 8611 WBs within 250 pc of the Sun. Orbits are integrated in a rigorously calculated gravitational field that directly includes the EFE. We also allow line-of-sight contamination and undetected close binary companions to the stars in each WB. We interpolate between the Newtonian and Milgromian predictions using the parameter α_{grav} , with 0 indicating Newtonian gravity and 1 indicating MOND. Directly comparing the best Newtonian and Milgromian models reveals that Newtonian dynamics is preferred at 19σ confidence. Using a complementary Markov Chain Monte Carlo analysis, we find that $\alpha_{\text{grav}} = -0.021_{-0.045}^{+0.065}$, which is fully consistent with Newtonian gravity but excludes MOND at 16σ confidence. This is in line with the similar result of Pittordis and Sutherland using a somewhat different sample selection and less thoroughly explored population model. We show that although our best-fitting model does not fully reproduce the observations, an overwhelmingly strong preference for Newtonian gravity remains in a considerable range of variations to our analysis. Adapting the MOND interpolating function to explain this result would cause tension with rotation curve constraints. We discuss the broader implications of our results in light of other works, concluding that MOND must be substantially modified on small scales to account for local WBs.

Key words: gravitation – methods: statistical – celestial mechanics – binaries: general – stars: kinematics and dynamics – galaxies: kinematics and dynamics.

1 INTRODUCTION

Our current best understanding of gravity is encapsulated by the theory of General Relativity (GR; Einstein 1915). This reduces to Newtonian dynamics in the weak-field non-relativistic limit (Rowland 2015; de Almeida, Piattella & Rodrigues 2016; Ciotti 2022). According to this theory, the gravity from a point mass follows an inverse square law, which historically was inferred from the fact that the rotation velocity v_c of a planet around the Sun declines with its heliocentric distance r as $v_c \propto 1/\sqrt{r}$. In principle, this Keplerian decline should also apply to galaxies as a whole if one considers tracers of their potential far beyond the bulk of their detectable baryonic mass (stars and gas). However, it is well known that galaxy rotation curves are flat (Bosma 1978; Rubin, Ford & Thonnard 1978; Faber & Gallagher 1979, and references therein) – this is nicely illustrated in fig. 15 of Famaey & McGaugh (2012).

Several explanations have been put forward for this missing gravity problem, which is also apparent in a number of other ways like dwarf galaxy velocity dispersions (fig. 11 of McConnachie 2012) and weak gravitational lensing (Brouwer et al. 2021). The most popular idea is that the extra gravity needed to bind galaxies and enhance their effect on passing photons comes from an otherwise undetected halo of particles beyond the well-tested standard model of particle physics (for a review, see Peebles 2017). This forms the basis for the currently prevailing Lambda cold dark matter (Λ CDM) standard model of cosmology (Efstathiou, Sutherland & Maddox 1990; Ostriker & Steinhardt 1995). However, DM particles have not been found despite decades of highly sensitive searches for collisions with terrestrial detectors (LUX Collaboration 2017; Lux-Zeplin Collaboration 2023). Similarly, there is no sign of γ -rays from DM annihilation or decay processes in supposedly DM-dominated Galactic satellites, providing stringent limits on the allowed properties of the hypothetical particles (Hoof, Geringer-Sameth & Trotta 2020).

This has led some workers to consider that the missing gravity problem might not be due to seemingly undetectable particles. The

* E-mail: indranilbanik1992@gmail.com (IB); w.j.sutherland@qmul.ac.uk (WS)

most developed alternative is Milgromian dynamics (MOND; Milgrom 1983). It proposes that the gravity g in an isolated spherically symmetric system is asymptotically related to the Newtonian gravity g_N of the baryons alone according to

$$g \rightarrow \begin{cases} g_N, & \text{if } g_N \gg a_0, \\ \sqrt{a_0 g_N}, & \text{if } g_N \ll a_0. \end{cases} \quad (1)$$

MOND introduces a fundamental acceleration scale a_0 below which the gravitational field deviates substantially from Newtonian expectations.¹ Just like with the Newtonian constant G , we must determine a_0 empirically. Using the rotation curves of a handful of disc galaxies with properties similar to our own, it has been found that $a_0 = 1.2 \times 10^{-10} \text{ m s}^{-2}$ (Begeman, Broeils & Sanders 1991). This value has remained very stable over many decades (Gentile, Famaey & de Blok 2011; McGaugh, Lelli & Schombert 2016; Chae 2022).

MOND works remarkably well at predicting the dynamics of galaxies across a huge range in baryonic mass, surface brightness, and gas fraction (Famaey & McGaugh 2012; McGaugh, Lelli & Schombert 2016; Lelli et al. 2017b; Kroupa et al. 2018; Li et al. 2018). Observations show that galaxies follow a remarkably tight ‘radial acceleration relation’ (RAR) between the radial components of g and g_N as deduced from rotation curves and the baryonic mass distribution, respectively (McGaugh 2020). This conclusion is not much dependent on assumptions about the stellar mass-to-light ratio because one can restrict attention to gas-rich galaxies (McGaugh 2012). The intrinsic scatter in the RAR is only 0.034 dex, which given the uncertainties is consistent with zero (Desmond 2023). The RAR also extends to elliptical galaxies and even to weak gravitational lensing down to $g_N \approx 10^{-5} a_0$ (Brouwer et al. 2021). The gravitational forces in virialized galaxies are thus very well described by MOND (for a review, see section 3 of Banik & Zhao 2022).

These successes are complemented by recent numerical MOND simulations of galaxies, which allow an exploration of non-axisymmetric dynamical features like bars and spiral arms (Banik et al. 2020; Roshan et al. 2021a). MOND has also been shown to get about the right star formation rates in disc galaxies with an appropriate size and gas fraction for their stellar mass (Nagesh et al. 2023). While these works typically target isolated galaxies, MOND has also been applied to interacting galaxies, in particular to understand the Antennae (Renaud, Famaey & Kroupa 2016), the tidal stability of dwarf spheroidal galaxies in the gravitational field of the Fornax Cluster (Asencio et al. 2022), the asymmetric tidal tails of the globular cluster Palomar 5 (Thomas et al. 2018) and open star clusters in the Solar neighbourhood (Kroupa et al. 2022), and the formation of the Local Group (LG) satellite planes around the Milky Way (MW) and M31 from tidal debris expelled by a past flyby encounter between their discs around 8 Gyr ago (Zhao et al. 2013; Bílek et al. 2018; Banik et al. 2022).

Many of these results were motivated by tensions and inconsistencies within the Λ CDM paradigm. For instance, it is known to face severe difficulties explaining the LG satellite galaxy planes (Pawlowski 2021b, and references therein). Hydrodynamical MOND simulations of M33 (Banik et al. 2020) were motivated by difficulties in understanding why it is weakly barred, bulgeless, and has a two-armed spiral – behaviour which is in many ways the opposite of what arises in Newtonian simulations with a live halo (Sellwood, Shen & Li 2019). Problems understanding galaxy bars in Λ CDM

are certainly not confined to M33: the model faces a 13σ tension when confronted with the observed ratios between bar lengths and corotation radii (Roshan et al. 2021b). This fast bar tension has been rather directly linked to dynamical friction between the rotating bar and the hypothetical CDM halo (fig. 12 of Roshan et al. 2021a). Those authors demonstrated that the problem is alleviated in MOND (see their fig. 20), with a similar result found in subsequent more detailed simulations that include hydrodynamics and star formation (see fig. 16 of Nagesh et al. 2023). The presence of bars in galaxies with a low surface brightness is also very problematic for Λ CDM because these galaxies are supposedly embedded in a dominant dark halo, which reduces the role of disc self-gravity (Kashfi, Roshan & Famaey 2023). Milgromian disc galaxies are always completely self-gravitating, making it easier to understand why even galaxies with a large enhancement to their Newtonian baryonic rotation curve commonly have bars and spiral arms (McGaugh, Schombert & Bothun 1995; McGaugh & de Blok 1998a, b).

MOND has also been applied on galaxy cluster and cosmological scales with the neutrino hot dark matter (ν HDM) paradigm developed by Angus (2009) and the aether scalar tensor (AEST) model (Skordis & Złotońik 2019, 2021, 2022) – for a review, we refer the reader to section 9 of Banik & Zhao (2022). Here again some of the successes relate to situations which are severely problematic for Λ CDM, in particular the KBC void and Hubble tension (Keenan, Barger & Cowie 2013; Haslbauer, Banik & Kroupa 2020) and the El Gordo massive galaxy cluster collision at redshift $z = 0.87$ (Asencio, Banik & Kroupa 2021, 2023). Thus, the successes of MOND extend well beyond its initial motivation forty years ago from the rotation curves of high surface brightness disc galaxies. At the same time, the motivation for considering alternatives to Λ CDM extends well beyond a mere failure to detect the hypothetical particles (for an extensive review, see Banik & Zhao 2022).

1.1 The importance of wide binaries (WBs)

In this contribution, we seek to test the validity of MOND on very small length scales by galactic standards, greatly diminishing the role of DM (Acedo 2020). One might think that the role of MOND would also be diminished, but since MOND does not introduce a fundamental new length scale, it should remain valid in systems with a much smaller mass and size than a typical galaxy provided the accelerations are low enough. Equation 1 indicates that Newtonian and Milgromian gravity diverge when $g \lesssim a_0$. Around an isolated point mass M , this occurs beyond its MOND radius

$$r_M \equiv \sqrt{\frac{GM}{a_0}}. \quad (2)$$

The MOND radius of the Sun is 7000 AU (7 kAU) or 0.03 pc. This is much smaller than the typical separation between stars in the Solar neighbourhood, where the Newtonian Jacobi or tidal radius for a system with $M = 2 M_\odot$ is 350 kAU (equation 43 of Jiang & Tremaine 2010). It is thus possible to test MOND at kAU distances from a star provided suitable tracers can be found. The properties of long-period comets that reach kAU distances have been used to argue in favour of the Milgromian enhancement to the Solar potential (Penner 2020), while its predicted anisotropy may explain the apparent orbital alignment of Kuiper belt objects with a large pericentre (Brown & Mathur 2023; Migaszewski 2023). However, this evidence is rather indirect. More direct evidence could be obtained using interstellar precursor missions travelling at a few percent of the speed of light

¹This is true in percentage terms, but the difference in absolute terms is $\lesssim a_0$.

c (Banik & Kroupa 2019b), but this remains well beyond current technology.

These difficulties can be overcome by using stars in WBs, defined in this work as binary stars with orbital acceleration $\lesssim a_0$, implying kAU-scale separations. The nearest star to the Sun is in such a WB (Beech 2009, 2011). MOND can thus be tested using the orbital acceleration of Proxima Centauri around α Centauri A and B, which have a very small separation and can be treated as one object (Kervella, Thévenin & Lovis 2017). Unfortunately, the $\approx 0.5 \mu\text{s}$ astrometric precision required is roughly two orders of magnitude beyond the reach of current instruments (Banik & Kroupa 2019a). Instead of using the acceleration, it is much more promising to use the relative velocity, but then the wide binary test (WBT) of gravity has to be done statistically by considering a large number of WBs to average over orbital phases and projection effects. This was first suggested by Hernandez, Jiménez & Allen (2012), with Pittordis & Sutherland (2018) later exploring the behaviour of WBs in several gravity theories. Since WBs come in a range of masses and separations, the WBT is usually phrased as a statistical test involving the distribution of the dimensionless parameter

$$\tilde{v} \equiv v_{\text{rel}} \div \overbrace{\sqrt{\frac{GM}{r_{\text{sky}}}}}^{\text{Newtonian } v_c}, \quad (3)$$

where M is the total mass of a WB with sky-projected separation r_{sky} and relative velocity v_{rel} , with $v \equiv |\mathbf{v}|$ for any vector \mathbf{v} . Thus, the relative velocity is normalized to the Newtonian circular velocity at the projected separation. Since the actual separation is larger, \tilde{v} is smaller than if the Newtonian v_c had been calculated using the 3D separation. Moreover, we only consider the sky-projected part of v_{rel} (which we denote \mathbf{v}_{sky}) as the data is not yet accurate enough for a 3D version of the WBT (though see Section 5.5). This is mainly due to the parallaxes being too inaccurate to reliably constrain the 3D separation (Section 2.3.2), though there is also some additional uncertainty from the required correction for the gravitational redshift of each star (Loeb 2022). Considering only sky-projected quantities further reduces \tilde{v} , which in this contribution is based on using \mathbf{v}_{sky} in equation (3) – we do not use the full 3D \tilde{v} . Note that even though we will mostly be dealing with the magnitudes of \mathbf{r}_{sky} and \mathbf{v}_{sky} in this work, these are actually 2D vectors within the sky plane. This is important to the work of Hwang, Ting & Zakamska (2022), who used the angle ψ between \mathbf{r}_{sky} and \mathbf{v}_{sky} to obtain interesting constraints on the WB orbital eccentricity distribution (Section 3.1.2).

The first detailed MOND calculations of WB orbits showed that the Milgromian circular velocity would be enhanced by 20 per cent over the Newtonian expectation in the asymptotic regime of large separations, a result which can be understood analytically (Banik & Zhao 2018c, hereafter BZ18).² Numerical MOND simulations of nearby star clusters also show a 20 per cent enhanced velocity dispersion over their Newtonian counterparts for essentially the same reason (Kroupa et al. 2022). The enhancement is limited by the Galactic gravity on the Solar neighbourhood, which is important in MOND because it is a non-linear theory [equation (1)]. As a result, the internal behaviour of a self-gravitating system is influenced by an external source of gravity, even in the absence of tidal effects. This non-standard external field effect (EFE) has been an important aspect of MOND since the beginning, not only because it is theoretically inevitable in a non-linear theory, but also because

the EFE is necessary to suppress the MOND enhancement to the velocity dispersions of the Pleiades and Praesepe open star clusters (see section 3 of Milgrom 1983).

Many further arguments for the EFE can be made nowadays (for a review, see section 3.3 of Banik & Zhao 2022). It significantly enhances the tidal susceptibility of dwarf galaxies in a cluster environment by weakening their self-gravity, thereby helping to explain the observed signs of tidal disturbance in the Fornax Cluster’s dwarf galaxies and the lack of low surface brightness dwarfs towards its centre (Asencio et al. 2022). As explained in that work, tidal radii are necessarily in the EFE-dominated regime, making their results particularly sensitive to the EFE. Its subtle imprint on the outer parts of disc galaxy rotation curves has recently been detected by comparing galaxies in isolated and more crowded environments (Chae et al. 2020b, 2021). The EFE is also crucial when calculating the escape velocity from a mass distribution, since the Milgromian potential of an isolated mass is infinitely deep (Famaey, Brunton & Zhao 2007; Zhao, Li & Bienaymé 2010; Banik & Zhao 2018b). This is not true once the EFE is included because at large distances where the EFE dominates, the gravitational field of a point mass returns to the Newtonian inverse square law, albeit with a higher normalization and some angular dependence (see fig. 1 of Banik & Zhao 2022). This EFE-dominated regime is sometimes known as the quasi-Newtonian regime, though in this contribution we simply refer to it as the asymptotic regime because local WBs lack an extended region in which they are isolated and at low acceleration. This is due to the Galactic external field g_e being slightly stronger than a_0 , as evident from the Galactic rotation curve (e.g. Zhu et al. 2023).

Given the importance of g_e for the WBT, it is fortunate that we know g_e from a direct measurement of the acceleration of the Solar System relative to distant quasars using their changing aberration angle, which directly tells us that $g_e/c = 5.05 \pm 0.35 \mu\text{s/yr}$ in a direction very close to that of the Galactic centre (Gaia Collaboration 2021b). This precludes substantial deviations of g_e from the value deduced kinematically from the Galactic rotation curve (Section 3.1). Even so, the predicted 20 per cent enhancement to the orbital velocities of local WBs is related to the shape of the MOND interpolating function in the regime close to the critical acceleration scale a_0 . It is thus somewhat model-dependent as it does not directly relate to the weak-field asymptotic limit at the heart of MOND. Nevertheless, recent detailed rotation curve studies have mostly converged on the form of the interpolating function in this regime, though of course slight variations are still possible. We will return to this point in Section 5.3.

Prospects for the WBT were discussed extensively in BZ18, whose section 8 clarified the main systematic effects that would likely have to be considered. The main issue was expected to be undetected close binary (CB) companions to one or both of the stars in a WB. Indeed, the nearest star to the Sun is in just such a triple system: Proxima Centauri is on a wide orbit about α Centauri A and B, whose mutual orbital semi-major axis of 23 AU is negligible compared to the 13 kAU distance from their barycentre to Proxima Centauri (Kervella et al. 2016; Kervella, Thévenin & Lovis 2017). At larger distances from the Sun, it is possible that a similar CB would not be resolved, but it would still have significant effects on the observed kinematics that would not arise if Proxima Centauri were orbiting a single star (Section 3.2.1). Pittordis & Sutherland (2019, hereafter PS19) considered the WBT in light of actual data from Gaia Data Release 2 (DR2; Gaia Collaboration 2018), highlighting that there is indeed an extended tail towards much higher \tilde{v} than could plausibly arise from genuine WB orbital motion, which cannot realistically yield $\tilde{v} \gtrsim 2$ for any reasonable modification to gravity.

² \tilde{v} in this contribution has the same meaning as \tilde{v}_{sky} in BZ18.

The possible presence of CBs was strongly suggested by the analysis of Clarke (2020) and later by Belokurov et al. (2020), who argued that some of the WBs with unexpectedly large relative velocities also have a *Gaia* astrometric solution that poorly fits the observations. Since the astrometric solution only includes parallax and proper motion, it was argued that the excess ‘noise’ could be due to astrometric acceleration induced by a CB, which was previously shown to typically induce a much larger orbital acceleration than the WB orbital motion (see section 8.2 of BZ18). Further follow-up observations should be able to confirm the existence of CB companions around some WBs (Manchanda, Sutherland & Pittordis 2023), especially once the astrometric time series are published in *Gaia* DR4 (see Section 5.5).

Some line-of-sight (LOS) contamination is also inevitable, though this is expected to be rather modest thanks to the excellent quality of the *Gaia* data (see section 3 of PS19). Their work shows a clear main peak to the \tilde{v} distribution due to WB orbital motion. Investigation of its properties already decisively rejects theories where local WBs behave like isolated MOND systems without the weakening of their self-gravity due to the Galactic EFE (see their fig. 11). This is because the gravity binding WBs can be deduced from the main peak to the \tilde{v} distribution at $\tilde{v} \lesssim 2$, with CB and LOS contamination leading to an extended low amplitude tail out to much higher \tilde{v} . If the orbital velocity becomes asymptotically flat with increasing separation, we would expect the mode of the \tilde{v} distribution to be significantly higher in systems with large r_{sky} . The predicted peak shift is in catastrophic disagreement with the observations, which show at most only a modest peak shift. This still leaves open the possibility that WBs obey MOND with its inevitable EFE, but it falsifies the quantized inertia proposal (McCulloch & Lucio 2019) and repeated claims to have confirmed that WBs follow the isolated MOND prediction without considering the full \tilde{v} distribution (Hernandez et al. 2019; Hernandez, Cookson & Cortés 2022).

In this contribution, we conduct a detailed statistical analysis to test the MOND prediction for local WBs. We use a rigorous grid solution to the MOND field equation including the EFE (BZ18) and then add possible perturbations from undetected CBs around one or both of the stars in each WB. We jointly infer properties of the WB population, the undetected CB population, and the extent of LOS contamination, which becomes important towards large r_{sky} and \tilde{v} . Both the model used and the exploration of its parameter space are much more detailed than in any previous attempt at the WBT. To mitigate possible biases, the plan was prepared in advance (Banik, Pittordis & Sutherland 2021) and has barely been modified to deal with the actual *Gaia* DR3 (Gaia Collaboration 2023).

In the following, we explain how we reduce the *Gaia* data set to a form suitable for our analysis and what quality cuts we employ (Section 2). We then introduce our detailed model for the WB data set and explain how we compare it to observations (Section 3). Our results are presented in Section 4 and discussed in Section 5, which includes a comparison with prior WBT results (Section 5.2). We conclude in Section 6.

2 THE OBSERVED WB DISTRIBUTION

The primary data set for the WBT comes from the precise results obtained by the *Gaia* mission (Gaia Collaboration 2016). Local WBs were extracted from the *Gaia* data set using the methods described in Pittordis & Sutherland (2023, hereafter PS23), who slightly adapted the methods in an earlier work (section 2 of PS19) based on *Gaia* DR2 (Gaia Collaboration 2018) due to the improved quality in *Gaia* Early Data Release 3 (EDR3; Gaia Collaboration 2021a). We supplement

EDR3 with radial velocities (RVs) from the full *Gaia* DR3 (Gaia Collaboration 2023).

In Section 2.1, we describe the basic quality cuts that we impose. We then conduct a Monte Carlo propagation of the uncertainties (Section 2.3). This leads us to impose further quality cuts, which we briefly summarize in Section 2.4 alongside our choice of restrictions on the parameter ranges.

2.1 The *Gaia* DR3 sample and basic quality cuts

The details of the sample selection are given in PS23, whose main points we summarize below. We begin by selecting all stars from *Gaia* EDR3 at a Galactic latitude $|b| > 15^\circ$ with an apparent *Gaia*-band (*G*-band) magnitude $m_G < 17$ and measured parallax $\varpi > 4$ mas (estimated distance < 250 pc) uncorrected for parallax bias (Lindegren et al. 2021). From this sample of 2.1 million single stars, candidate WBs are selected by requiring:

- (i) Projected separation $r_{\text{sky}} \leq 50$ kAU;
- (ii) Star distances consistent with each other within the lesser of 8 pc or $4 \times$ the combined distance uncertainty, i.e. $|d_A - d_B| \leq \min(4\sigma_d, 8 \text{ pc})$, where *A* and *B* label the stars in each candidate WB; and
- (iii) Projected velocity difference between the stars of $v_{\text{sky}} < 3 \text{ km s}^{-1}$, as inferred from the difference in proper motions but assuming both stars are at the mean estimated distance $\bar{d} \equiv (d_A + d_B)/2$.

From this preliminary list of binaries (WB-EDR3), sky regions are removed around four known open clusters (see table 1 of PS23).

To remove some probable triple systems or groups, we reject all WBs in which either star is common to more than one candidate binary in WB-EDR3. We also search for comoving companion stars to a fainter limit using a ‘faint star’ sample constructed from *Gaia* EDR3 stars with $m_G < 20$ and parallax $\varpi > 10/3$ mas. For each star in each candidate WB, we search for companions in the faint star sample with the following criteria:

- (i) Parallax distance consistent with the main star at 4σ ;
- (ii) Angular separation less than $2/3$ of the main binary separation (since hierarchical triples are expected to be unstable if the inner orbit separation $\gtrsim 0.4 \times$ the outer separation);
- (iii) Angular separation > 0.5 arcsec to avoid barely-resolved companions; and
- (iv) Projected velocity difference from the main star $\leq 5 \text{ km s}^{-1}$.

If any such ‘third star’ is found, the candidate binary is rejected. Note however that this will not reject triples with an inner orbit closer than about 100 AU or where the third star is so faint that $m_G > 20$, so it remains crucial to model the CB population. We return to this point in Section 3.2.3.

PS23 also applied a per-star quality cut based on the *Gaia* parameters (see equation 1 of Arenou et al. 2018). WBs were only considered if both stars satisfy

$$\sqrt{\frac{\chi^2}{\nu}} \leq 1.2 \max(1, \exp[-0.2(m_G - 19.5)]), \quad (4)$$

$$\chi^2 \equiv \text{astrometric_chi2_al},$$

$$\nu \equiv \text{astrometric_n_good_obs_al} - 5.$$

This cut is based on *Gaia* DR2, so it should be fairly conservative in *Gaia* DR3 given the greater number of observations per star and the longer time baseline. Throughout this study, any quality cuts applied at the individual star level are implemented by excluding the whole WB if either of its two stars fails to pass the relevant quality cut.

Applying all the above quality cuts yields a sample of 73k candidate WBs. The majority have $r_{\text{sky}} \leq 1$ kAU, which is not very useful for the WBT. We therefore publicly release a smaller version of this sample consisting of 19 786 WBs that satisfy $1.5 < r_{\text{sky}}/\text{kAU} < 40$, which is somewhat wider than the parameter range we use for the WBT.

From the *Gaia* DR3 WBs that pass the above quality cuts, we restrict ourselves to the range $r_{\text{sky}} = 2\text{--}30$ kAU and $\tilde{v} \leq 5$. These choices were fixed in advance to limit the possibility of biasing the results (Banik, Pittordis & Sutherland 2021). The upper limit to \tilde{v} is much larger than can plausibly arise from WB orbital motion. This is necessary to properly sample the extended tail to the \tilde{v} distribution, which we need in order to properly constrain the CB and LOS contamination causing the tail. In particular, LOS contamination is expected to become more significant at high \tilde{v} because \tilde{v} is a 2D quantity, so we can get much better leverage on the extent of LOS contamination if we have data at high \tilde{v} . Even so, we do not want to include data at very high \tilde{v} because the likelihood of systematics increases the wider the ‘aperture’ used for the WBT.

We could increase the sample size substantially by going down to even lower separations, but then there would be a risk of swamping the analysis with data on Newtonian systems. In this case, small improvements to the fit in the Newtonian regime would become more important to the statistical analysis than checking whether the \tilde{v} distribution differs between systems within their MOND radius and those with larger separations. On the other hand, we need to go down to rather low r_{sky} to provide a ‘Newtonian anchor’ population that cannot be affected by MOND, especially since the RAR requires a fairly smooth MOND interpolating function such that significant MOND effects persist even when the accelerations slightly exceed a_0 (Section 5.3). Model parameters besides the gravity law can be constrained much better if we have a significant number of WBs which cannot be much affected by MOND, since then the parameters inferred here will not be degenerate with the gravity law. In particular, a Newtonian anchor population should be extremely valuable in constraining the CB population parameters – WBs are expected to have similar CB populations regardless of whether their WB orbital acceleration is below or above a_0 . Our analysis does not explicitly consider some subset of our sample to be completely immune to MOND (unlike Chae 2023), but instead relies on the predicted MOND effects being different at low and high r_{sky} . We will see later that since our sample goes up to $M = 4 M_\odot$ and thus has systems with MOND radii of up to 14 kAU [equation (2)], we have plenty of WBs separated by a small fraction of their MOND radius. This is especially true given the steeply declining r_{sky} distribution of WBs (BZ18, PS19, PS23), which also implies that the 3D separation is likely to only slightly exceed r_{sky} [equation (8)].

2.2 The mass–luminosity relation

We estimate the mass of each star in our WB sample from its absolute *G*-band magnitude M_G using a similar technique to PS19, which we briefly describe below. We use the M_V and $(V - I)$ colour at different stellar masses M as tabulated in Pecaut & Mamajek (2013).³ We relate this to *Gaia* photometry using the relation between $(V - I)$ and $(G - V)$ colours from the first Johnson-Cousins relation in table C2

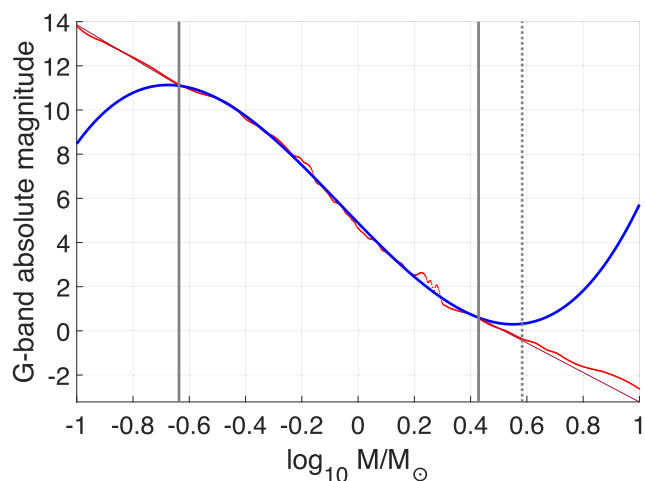


Figure 1. The relation between the mass of a star and its absolute *G*-band magnitude (red line). Results are based on the absolute *V*-band magnitude and $(V - I)$ colour tabulated in Pecaut & Mamajek (2013) and the relation between $(V - I)$ and $(G - V)$ colours from the first Johnson-Cousins relation in table C2 of Riello et al. (2021). Our cubic fit to the data [blue line; equation (6)] is used over the mass range between the vertical solid grey lines. Outside this range, a linear relation is assumed, as indicated by the solid purple lines. The dotted vertical line shows the maximum mass of any star in our WB sample.

of Riello et al. (2021), which states that

$$(G - V) = -0.01597 - 0.02809(V - I) - 0.2483(V - I)^2 + 0.03656(V - I)^3 - 0.002939(V - I)^4. \quad (5)$$

We now have a complete set of equations to obtain M_G for any M . However, this is difficult to invert because of small-scale irregularities in the relation arising from the complexities of stellar astrophysics (Fig. 1). We therefore fit the relation using the cubic

$$M_G = 4.887 - 5.693x + 0.4164x^2 + 0.9611x^3, \quad (6)$$

where $x \equiv \ln M/M_\odot$. Since each unit decrease in M_G corresponds to slightly less than an e-fold increase in luminosity L , equation (6) tells us that $L \propto M^5$ when $M = M_\odot$ ($x = 0$), representing a steep mass–luminosity relation. We use our cubic fit over the range $0.6 < M_G < 11.1$. At the low (high) mass end, we assume the relation becomes linear and has a slope of -3.3 (-2.9). The fitted relation between M_G and x is shown as the solid blue line, confirming that the above cubic provides a fairly good fit over the range we use it, i.e. between the vertical solid grey lines. The purple lines outside this range show that a linear relation here also provides a reasonably good fit, at least up to the dotted vertical grey line showing the maximum mass of any star in our sample. A cubic is the lowest order polynomial which seems to provide a good match to the tabulated data in Pecaut & Mamajek (2013), mainly because the slope is similar at low and high masses but there is a steepening at intermediate masses.

We invert our piece-wise analytic relation between $\ln M/M_\odot$ and M_G to obtain the mass of any star from its M_G , which we obtain from its apparent magnitude using its trigonometric parallax. The linear relations assumed at very low and very high masses can be readily inverted. We invert our cubic fit over most of the mass and M_G range using the Newton–Raphson method. To speed this up, we obtain an initial guess for the mass using a linear fit of the form

$$M_G \approx 5.023 - 5.102x. \quad (7)$$

³https://www.pas.rochester.edu/~emamajek/EEM_dwarf_UBVIJHK_colors_Teff.txt [3.3.2021]

This can be inverted analytically without iterative techniques. Our estimate that $-\frac{d \ln M}{d M_G} = 0.20$ is similar to previous estimates of $0.074 \ln 10 = 0.17$ (equation 3 of PS23) and $0.0725 \ln 10 = 0.17$ (equation 3 of PS19).

We consider uncertainties in M_G arising from those in the trigonometric parallax and the *Gaia* photometric uncertainty in the apparent magnitude, which we convert from a fractional uncertainty to a magnitude uncertainty. We also allow a small uncertainty in the conversion of luminosity to mass. This must play some role, but the ≈ 42 per cent asymptotic enhancement to the radial gravity predicted in MOND combined with our estimate that $\frac{d M_G}{d \ln M} = -5.1$ [equation (7)] implies that the MOND signal corresponds to a difference in M_G of about $5.1 \ln 1.4 = 1.7$ mag, which is very large compared to the slight fluctuations about our cubic mass-luminosity relation evident in Fig. 1. In other words, the steep mass-luminosity relation implies that mass uncertainties should have a very small impact compared to uncertainties in the relative velocity arising from uncertain proper motions, especially when the key quantity entering \tilde{v} is essentially the relative proper motion between the two stars in each WB. The mass of a star with known absolute magnitude can be constrained to within ≈ 6 per cent (Eker et al. 2015; Mann et al. 2015), but due to the square root term in equation (3), this would have only a 3 per cent impact on \tilde{v} . The impact on the WBT is further reduced because it is mostly concerned with the width of the \tilde{v} distribution, which is only affected at second order by measurement uncertainties. Even so, we do assign a modest 0.024 dex or 5.5 per cent uncertainty to our stellar mass estimates at fixed luminosity for reasons discussed in Section 2.3.3. We will see later that our main conclusion is the same as a different study which used a linear relation between M_G and x over the full mass range relevant for the WBT, thus leading to rather different errors in the conversion of absolute magnitudes to masses (PS23).

2.3 Monte Carlo error propagation

We use the full 5×5 *Gaia* covariance matrix to propagate uncertainties in the sky position, parallax, and proper motion of each star in our WB sample into an uncertainty in \tilde{v} . We will see later that this uncertainty is very small for a large number of WBs, so our approach is to consider only those systems and obtain their $(r_{\text{sky}}, \tilde{v})$ directly from the raw observables. The resulting $(r_{\text{sky}}, \tilde{v})$ distribution is then used in our main analysis, which therefore does not directly include measurement uncertainties – these are included indirectly in the sample selection. Our procedure avoids adding measurement errors to values that already contain measurement errors, as that would lead to adding measurement errors twice to the latent values.

For our Monte Carlo error propagation, we diagonalize the correlation matrix and generate Gaussian random numbers to propagate uncertainties along each of the eigenvector directions, bearing in mind also the eigenvalues. The correlation matrix is dimensionless, so we relate it to the covariance matrix using the *Gaia* catalogue uncertainty of each parameter. Uncertainties in the parallax and proper motion are important, but we do not propagate the uncertainty in the sky position of each star at the reference epoch, instead taking the published position at face value. This simplifies our analysis greatly because it enables us to work with a fixed definition of the sky plane for each WB, which we take to be the plane whose normal vector is the angular bisector of the directions toward each of the stars in the WB. Since the sky positions of the stars in our WB catalogue are expected to be particularly precise, there should be negligible uncertainty associated with this definition. Apart from this

minor change, we follow the same method as that described in Banik (2019) using mock parallax and proper motion data which samples the *Gaia* error distribution as described above. We assume that the projected separation of each WB has only a negligible uncertainty, so we obtain r_{sky} from the raw data.

2.3.1 The systemic RV

Compared to the detailed plan we set up earlier (Banik, Pittordis & Sutherland 2021), a major simplification has become possible with regards to the systemic RV of each WB because we can generally obtain this directly from *Gaia* DR3 (Gaia Collaboration 2023) without having to guess it from the proper motion and distance to the Galactic disc plane. The RV of each star enters the calculation of v_{sky} through perspective effects (Shaya & Olling 2011; Banik 2019; El-Badry 2019) – a receding WB with no internal motion will have a shrinking apparent separation on the sky. The impact is suppressed by the angular separation of the WB on our sky plane, but the systemic RV still cannot be completely neglected: for a WB with $r_{\text{sky}} = 10$ kAU, a typical systemic RV of 20 km s^{-1} translates into an apparent proper motion of 10 m s^{-1} at a heliocentric distance of 100 pc. Since the Newtonian v_c of two Sun-like stars with a 10 kAU separation is 420 m s^{-1} , the impact on \tilde{v} would be $\lesssim 0.03$ depending on the geometry. While this is small, we opt to try and include the impact of recessional motion because WBs can be even closer and/or have a larger RV.

We can see from the above that the at most km s^{-1} level difference in RV between the stars in each WB would have only a negligible impact on \tilde{v} . This is because relative motions along the LOS and within the sky plane should be comparable, but the impact of the former are suppressed by the angular WB separation on the sky. Consequently, we limit ourselves to including only the systemic RV of each WB. We find the systemic RV by taking an inverse variance weighted mean of each star’s RV when this is available for both stars, leading to an uncertainty smaller than either star’s RV uncertainty.⁴ If the RV is only available for one star, we adopt this as the systemic value. We then assign both stars a recessional velocity equal to the systemic RV, errors in which are propagated as part of the Monte Carlo error propagation.

2.3.2 Reducing uncertainty in the relative distance

Another source of perspective effects (discussed further in Shaya & Olling 2011) is the difference in heliocentric distances to the stars in each WB. The relative LOS separation is important because proper motions must be multiplied by distances to convert them from angular to physical velocities. The *Gaia* parallaxes typically do not provide very tight constraints on the LOS separation: while the astrometry is spectacular by historical standards, trigonometric distances to stars 100 pc away are not accurate at the kAU level (the typical uncertainty is about 80 kAU; see section 6.2 of BZ18). Using the same arguments as in Section 2.3.1, a system with a tangential motion of 20 km s^{-1} but whose LOS separation is uncertain by 80 kAU would have an almost 80 m s^{-1} uncertainty in v_{sky} . This translates to an uncertainty in \tilde{v} of up to 0.2 depending on the geometry, which could hamper the WBT.

⁴We combine the RV measurements when these are known for both stars because the typical RV uncertainty is around 1 km s^{-1} , which is much more than the expected WB relative velocity. Thus, differences in the RV are best understood as due to random errors rather than intrinsic WB orbital motion.

We intuitively expect that a WB's 3D separation is not much larger than its r_{sky} , as suggested by the referee to Banik (2019). We exploit this insight using the technique discussed in its section 2.3, which we briefly summarize here. The mean of the heliocentric distances to the two stars is taken to be the inverse variance weighted mean of the parallax distances to each of the stars in the WB, causing the mean distance to vary slightly between trials and thereby affect the distance modulus. Assuming a power-law prior on the 3D separation r of the form $P(r) \propto r^{-1.6}$ (Lépine & Bongiorno 2007; Andrews, Chanamé & Agüeros 2017; El-Badry & Rix 2018), we obtain a posterior inference on the ratio $x \equiv r/r_{\text{sky}}$. Following equation 18 of Banik (2019), we obtain that

$$P(x) \propto \frac{x^{-2.6}}{\sqrt{x^2 - 1}}, \quad x \geq 1. \quad (8)$$

As part of the Monte Carlo error propagation, we then randomly sample this distribution to obtain the separation of each WB along the LOS. We then push one of the stars away from us and bring the other star towards us by half this amount, with the displacement of each star being purely along its sky direction at the reference epoch. We use a random coin toss in each Monte Carlo trial to decide which star is to be pushed away from us. The revised distance to each star alters its estimated absolute magnitude and thus its mass.

This technique drastically reduces the uncertainty on the relative LOS distance in systems with a very low r_{sky} (compare the black and blue points in fig. 3 of Banik 2019). One disadvantage is that it implicitly assumes that the WB is genuine, so our results are less robust against LOS contamination. However, we mitigate this in our statistical analysis by including an allowance for LOS contamination (Section 3.3).

Our approach is similar to that used in Nelson et al. (2021), though they used a prior of $r^{-1.5}$ instead of $r^{-1.6}$. A more important difference is that those authors were only interested in the 3D separation between the stars, whereas we need to find the heliocentric velocity of each star, which requires knowledge of the individual LOS distances. This causes our calculated \tilde{v} parameter to depend on which star is closer to us. Since we do not have this information, we use Monte Carlo trials to quantify the resulting uncertainty. We stress that in our main analysis, the adopted value of \tilde{v} is based on assigning the same LOS distance to both stars. The purpose of our deprojection is to quantify the uncertainty and combine this with other sources of uncertainty so that we can remove systems with insufficiently precise \tilde{v} .

The technique discussed above assigns the relative LOS distance a value of zero and a non-Gaussian uncertainty of $\approx r_{\text{sky}}$. This would only reduce the uncertainties if the parallaxes were not accurate enough to directly constrain the relative distance. Thus, we only apply the above technique if r_{sky} is smaller than the uncertainty in the relative LOS distance given by the *Gaia* trigonometric parallaxes. While we generally expect this to be the case, we do not enforce it. The 3D geometry of nearby WBs may well be clear from existing *Gaia* data, with the definition of 'nearby' extending to larger distances as the data improves. Our approach is thus to use whichever method is expected to be more accurate. Note that if *Gaia* parallaxes are used directly, then the relative distance along the LOS may have a non-zero mean value when averaged across different Monte Carlo trials as these would propagate uncertainties in each parallax. In this case, the nominal values adopted for our main analysis do not rely on the deprojection algorithm outlined above because each star is assigned its observed parallax distance.

2.3.3 Improved mass estimates

It is possible to slightly improve the accuracy of our mass estimates by using the Final Luminosity, Age and Mass Estimator (FLAME; Pichon 2007) work package in *Gaia* DR3. Masses obtained in this way are not available for all stars, but when they are available, they should be much more precise because they involve a detailed analysis of the spectrum. Fig. 3 of Hernandez (2023) shows that FLAME masses can differ by $\approx 0.05 M_{\odot}$ from estimates using a linear relation between absolute magnitude and the logarithmic mass. Unfortunately, restricting to only WBs where both stars have a FLAME mass would reduce the sample size too much for the WBT to be feasible. We therefore apply a small correction to the masses estimated using our cubic fit to the Pecaut & Mamajek (2013) mass-luminosity relation [equation (6)]. Denoting these masses by M_{PM} and the revised masses with the FLAME calibration as M_F , a good fit is given by

$$M_F = M_{PM} - 0.07 \tanh\left(\frac{M_{PM} - 0.75 M_{\odot}}{0.16 M_{\odot}}\right) f, \quad (9)$$

$$f = \begin{cases} \exp\left(2 - \frac{M_{\odot}}{M_{PM}}\right), & \text{if } M_{PM} \leq 0.5 M_{\odot}, \\ 1, & \text{if } 0.5 M_{\odot} \leq M_{PM} \leq 0.75 M_{\odot}, \\ \exp\left(\frac{0.75 M_{\odot} - M_{PM}}{2 M_{\odot}}\right), & \text{if } M_{PM} \geq 0.75 M_{\odot}. \end{cases}$$

The adjustment is tapered using the factor f so it rapidly decays for stars with mass $M < 0.5 M_{\odot}$ as there are no FLAME masses below this, so we assume that M_{PM} becomes very reliable at the low-mass end. We will see later that our results are not much affected by restricting our WB sample to systems where both stars have $M > 0.5 M_{\odot}$, which has the advantage that FLAME results can serve as a calibration. M_{PM} and FLAME masses gradually converge when $M \gg 0.75 M_{\odot}$.

Our small adjustment to M_{PM} is designed to eliminate a small systematic discrepancy with the FLAME mass in the stars for which the latter is available. Even so, our calculated M_F does not exactly coincide with the FLAME mass. Based on the typical difference between the two mass estimates and assuming that the FLAME masses are very accurate, we assign our estimated masses a random Gaussian uncertainty of 0.024 dex or 5.5 per cent at fixed absolute magnitude. A 5.5 per cent uncertainty is in line with the fact that masses estimated directly from luminosities were expected to have a 6 per cent uncertainty (Eker et al. 2015; Mann et al. 2015), as argued in section 6.3 of BZ18. Note that our mass estimates also include uncertainty in the absolute magnitude due to that in the apparent magnitude and the trigonometric parallax.

2.4 Refined quality cuts for the WBT

The quality cuts discussed in Section 2.1 provide a reasonably carefully prepared sample of WBs that can be used in further analyses. Unfortunately, some of these WBs are unsuitable for our highly precise analysis. The additional quality cuts in this work beyond those mentioned there are to exclude:

- (i) Stars in sky directions with a high total Galactic extinction towards an extragalactic source (Section 2.4.1);
- (ii) Stars which are substantially below the main sequence on a colour–luminosity diagram, which can make the mass estimate inaccurate (Section 2.4.2);
- (iii) Stars which sometimes appear as multiple peaks in the *Gaia* images, suggesting the presence of a CB companion that can sometimes be resolved (Section 2.4.3);

- (iv) WBs where the RV is unknown for both stars (Section 2.4.4);
- (v) WBs where the RV is known for both stars and there is a mismatch at $>3\sigma$ confidence (Section 2.4.5); and
- (vi) WBs whose \bar{v} is too uncertain (Section 2.4.6).

We also restrict ourselves to WBs with $2 < r_{\text{sky}}/\text{kAU} < 30$ and $\bar{v} < 5$ for reasons discussed earlier. To reduce the computational cost, we only consider WBs with a total mass of $0.464 - 4.31 M_{\odot}$. These cuts are now described further.

2.4.1 Dust extinction

As discussed in Section 2.2, our estimated mass for each star relies entirely on its absolute *G*-band magnitude, which in turn is inferred from its apparent magnitude and trigonometric parallax. The difference between apparent and absolute magnitudes is assumed to arise purely from the distance modulus, which implicitly assumes negligible dust extinction. This should be a good assumption for stars within 250 pc and $>15^{\circ}$ from the Galactic disc, especially as known star clusters are excluded (table 1 of PS23). Even so, we can further reduce the impact of dust by considering only WBs where both stars have a *V*-band extinction $A_V < 0.5$ based on the dust maps of Schlegel, Finkbeiner & Davis (1998), which reduces our sample size by 8.5 per cent. We note that their published values refer to the total Galactic extinction along the LOS towards an extragalactic source. Since the WBs in our sample are all within 250 pc, the extinction towards any of the stars in our sample should be far smaller, making our cut on A_V quite conservative.

2.4.2 Colour–magnitude diagram

Our mass estimates also rely on a mass–luminosity relation designed for main sequence stars (Pecaut & Mamajek 2013). To check if this is a good assumption, we use Fig. 2 to plot the relation between the absolute *G*-band magnitude and the colour, which we define as the difference in apparent magnitude between the *Gaia* blue and red passbands. As expected, the vast majority of the stars in our sample are on the main sequence. However, a small number of stars are very far below the main sequence – these are probably white dwarfs (WDs). We remove these by excluding stars below the solid blue line, which achieves a good separation between the main sequence and the WD track. Only a very small number of WDs are removed by this cut and these are all well separated from the main sequence, so any remaining WDs should have a negligible impact on our results.

An interesting feature of our main WB sample’s colour–magnitude diagram is a concentration of stars parallel to but brighter than the main sequence (Fig. 2). This ‘double main sequence’ is caused by unresolved CBs where both stars have an exactly equal mass, as demonstrated by the parallel dashed grey lines with an offset corresponding to a factor of two in luminosity. We do not exclude stars belonging to this double main sequence because our analysis already accounts for it (Appendix A).

A major complication to the WBT is the presence of undetected CB companions (Section 1). We follow a ‘mitigate + simulate’ strategy to minimize the uncertainties introduced by CBs. This entails reducing the presence of CB companions and modelling the distribution of those that remain. Since we do not aim to have a sample that is completely free of stars with a CB companion, we benefit from a much larger sample size than the study of Hernandez (2023). We discuss below how we attempt to reduce the proportion of contaminated WBs. Details of our CB model will be presented later (Section 3.2).

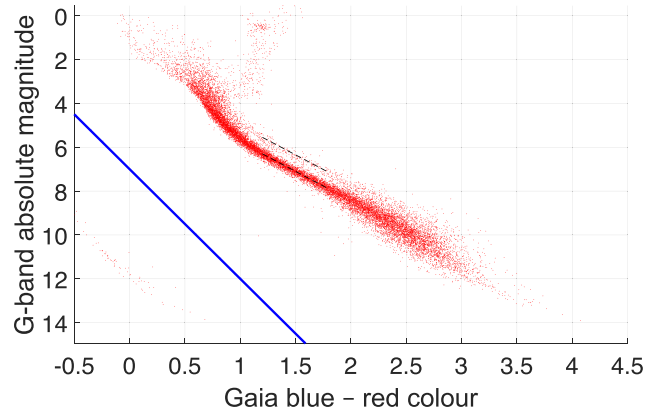


Figure 2. The colour–magnitude diagram of our sample just prior to removing 66 WBs where one star is below the solid blue line, which is designed to remove the white dwarfs evident at the lower left. The two dashed grey lines have an offset corresponding to a factor of two in luminosity. The lower line passes through the main sequence, while the upper line passes through the secondary track evident at higher luminosity. We attribute this double main sequence to unresolved binaries where the stars have an exactly equal mass (Appendix A).

2.4.3 Multiple peaks in *Gaia* images

If there is a clearly detected CB companion to a star in a WB and it is reasonably likely that the third star is bound to the WB, then the system is excluded from our analysis (Section 2.1). However, CBs which are fainter and/or on a tighter orbit might not be clearly resolved. The contaminated star might then appear resolved into two sources in only some *Gaia* focal plane transits. We exploit this by using the parameter *ipd_frac_multi_peak* in the *Gaia* DR3 catalogue (see also section 5 of El-Badry, Rix & Heintz 2021). This is the percentage of successful focal plane transits in which a source is detected as multiple peaks in the image. A genuinely isolated star would ideally give *ipd_frac_multi_peak* = 0, but since it might spuriously show up as multiple peaks in the *Gaia* images due to issues like cosmic ray hits, we allow *ipd_frac_multi_peak* ≤ 2 (a threshold of 2 was also used in Pace, Erkal & Li 2022). We reject any star that is detected as multiple peaks more frequently.

Fig. 3 shows the distribution of *ipd_frac_multi_peak* for the WBs in our sample prior to imposing this quality cut, with the higher value shown for the two stars in each WB. It is apparent that our quality cut does not reduce the sample size very much. At the same time, it seems unlikely that a star really has a marginally detected companion if it appears to be isolated 98 per cent of the time. However, if a star frequently appears as multiple peaks in the *Gaia* images, then we might reasonably suspect it to be part of a CB.

2.4.4 Having a systemic RV

Even though we are implementing a 2D version of the WBT using only motions within the sky plane, the systemic RV of each WB does still enter into our analysis, albeit scaled down substantially by the angular separation of the WB (Section 2.3.1). For our technique to work, we are forced to reject WBs where the RV is unknown for both stars, which loses roughly 1/3 of our sample. We accept systems where the RV is known for only one star because the systemic RV is not needed very precisely and the RVs of the stars in a WB would differ by at most a few km s^{-1} , especially for the more widely separated systems where perspective effects are more important. Requiring both stars to have a measured RV reduces the sample

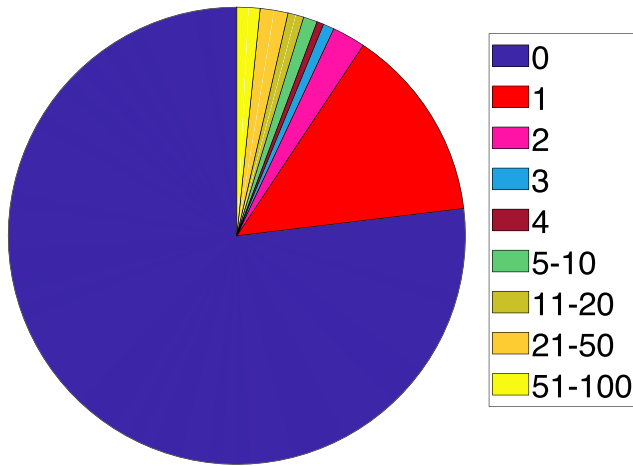


Figure 3. Distribution of the *Gaia* DR3 parameter *ipd_frac_multi_peak* for the WBs in our sample prior to imposing the quality cut related to this parameter. For each WB, we take the star with the higher value. Although *ipd_frac_multi_peak* can be as high as 100, it is typically 0, as expected for a single source.

size from 8611 to 5504, but the proportion of systems with $\tilde{v} > 2.5$ (which is a rough proxy for the level of contamination) only drops from 10.2 per cent to 9.5 per cent.

2.4.5 Similar RVs when both are available

While we do not require both stars in a WB to have a known RV, this is often the case. When available, we use this information to reject WBs whose two stars have RVs that differ from each other by more than triple the quadrature sum of the RV uncertainties or triple the Newtonian v_c , whichever is larger.⁵ This reduces the sample size by 3.5 per cent (we lose 308 WBs out of 8919 without this cut). Since almost 2/3 of the systems in our WB sample have a reported RV for both stars, the fact that only a small proportion of WBs consist of stars with discrepant RVs suggests that our selected WBs are mostly genuine. The use of systemic RVs constrained to km s^{-1} precision by *Gaia* spectroscopy substantially reduces the scope for perspective effects to alter the results obtained in this contribution.

2.4.6 The \tilde{v} uncertainty

The above quality cuts are designed to reduce systematic uncertainties in the WB parameters. These also have random uncertainties, which we need to quantify. As discussed in Section 2.3, we use Monte Carlo error propagation to obtain the uncertainty in the all-important \tilde{v} using 2^{12} trials. We find that the uncertainty is very small for a large proportion of the WBs in our sample (as expected from a simpler analysis of *Gaia* DR2 results; Banik 2019). We therefore apply a quality cut such that

$$\text{Error in } \tilde{v} \leq 0.1 \max\left(1, \frac{\tilde{v}}{2}\right). \quad (10)$$

We found that this achieves a good balance between the quality and quantity of data. The uncertainty is allowed to be somewhat larger at higher \tilde{v} because we use wider bins in \tilde{v} here when comparing

⁵Even if we had perfect data, the RVs of the stars in a WB could still differ by order their Newtonian v_c .

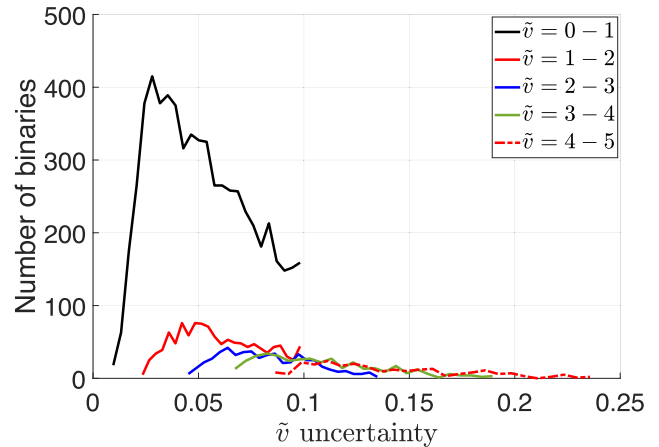


Figure 4. The frequency distribution of the \tilde{v} uncertainty as estimated from Monte Carlo error propagation of the 5×5 *Gaia* DR3 covariance matrix (Section 2.3). Different curves show results for different ranges of \tilde{v} , as indicated in the legend. We only consider WBs whose \tilde{v} uncertainty is at most $0.1 \max(1, \tilde{v}/2)$, truncating each error distribution at the right. The truncation at the left is caused by the 0.024 dex mass uncertainty at fixed M_G imposing a minimum fractional uncertainty in \tilde{v} .

to theory (Section 3.5). Moreover, such systems are less crucial for the WBT as they are not part of the main peak at $\tilde{v} \lesssim 2$, implying significant contamination from some source.

The distribution of \tilde{v} uncertainties is shown in Fig. 4 for different \tilde{v} ranges of our WB sample. As the WBT is mostly concerned with the main peak to the \tilde{v} distribution at $\tilde{v} \lesssim 2$, our cut implies a maximum \tilde{v} uncertainty of only 0.1 in the most crucial parameter range. Since the mode of the main peak is at $\tilde{v} \approx 0.5$, measurement uncertainties that add an extra 0.1 in quadrature would only boost the root mean square value by about 2 per cent, which is negligible in comparison to the predicted MOND enhancement of 20 per cent. Moreover, the typical \tilde{v} uncertainty for such systems is much smaller than the maximum permitted by our selection. Indeed, PS19 found the \tilde{v} distribution in *Gaia* DR2 changed little when restricting to systems where the uncertainty is below 0.25 (compare their figs 12 and 13). We therefore neglect \tilde{v} uncertainties in the rest of our analysis, bearing in mind that our WB sample has already been chosen to have an accurate \tilde{v} .

We need to integrate WB orbits for a range of masses covering the mass range of our sample. This is because larger mass systems have a larger MOND radius [equation (2)], reducing the impact of MOND at fixed separation (BZ18). To limit the computational cost, we restrict our analysis to WBs with a total mass of $0.464 - 4.31 M_\odot$. This reduces our sample size very slightly. Including the remaining WBs would require a considerable broadening of the mass range covered by our WB orbit library or a substantial worsening of its mass resolution, but there would be very little gain in sample size. It would also leave our results more vulnerable to systematic issues with very low or very high mass stars.

2.5 The WB sample

After applying the quality cuts described so far, we are left with a final sample of 8611 WBs. Their mass distribution is shown in Fig. 5, with blue bars showing individual stars and red bars showing binary total masses. The latter are more relevant physically, so we will use M to mean the binary total mass and M_* to denote the mass of a single star – unless it is clear that the discussion refers to stars considered individually.

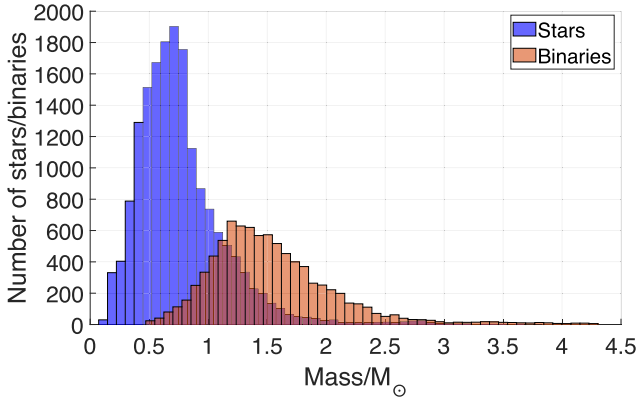


Figure 5. The mass distribution of all the stars in our WB sample (blue bars). The red bars show the distribution of the total mass of each binary. The outer edges of the outermost bins show the range of the data.

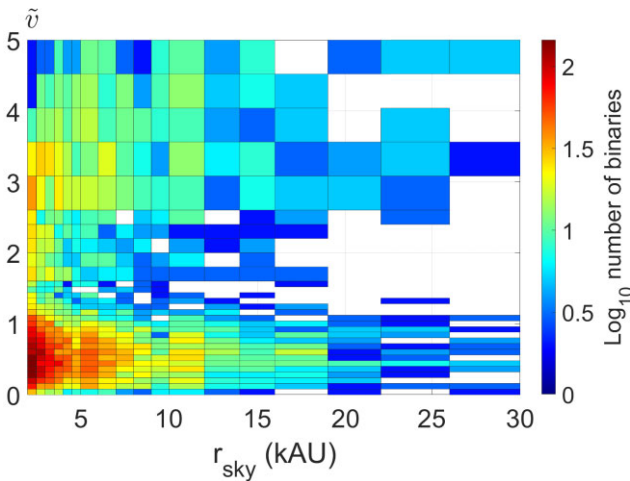


Figure 6. The number of WBs in different pixels of $(r_{\text{sky}}, \tilde{v})$. Results are shown binned according to our pixellation scheme, which was fixed prior to any detailed analysis (the pixel widths increase gradually towards higher values). The comparison with theory is based on the likelihood of matching the number counts shown here, which we find by multiplying together the binomial likelihood for each pixel (Section 3.5). Notice the lack of WBs at $\tilde{v} \approx 2$ for the highest r_{sky} (empty pixels are shown in white). We interpret this as due to dilution of the CB contamination over a wide \tilde{v} range due to the low Newtonian v_c of the WB. There is also little LOS contamination here: it only becomes important when r_{sky} and \tilde{v} are both large (Section 3.3).

As with previous detailed investigations into the WBT (PS19, PS23), we perform it by decomposing the data into pixels in the space of r_{sky} and \tilde{v} (Fig. 6). It is clear that there are many more systems at low r_{sky} than at high r_{sky} , in line with earlier results (e.g. Andrews, Chanamé & Agüeros 2017). The main peak to the \tilde{v} distribution is quite prominent, with the mode located at $\tilde{v} \approx 0.5$. There is also an extended tail going out to much larger \tilde{v} , which was previously reported (PS19) and attributed to undetected CBs (Belokurov et al. 2020; Clarke 2020). Interestingly, a gap appears in the distribution at $\tilde{v} \approx 2$, but only at high r_{sky} . We attribute this to the minimum in the \tilde{v} distribution at $\tilde{v} \approx 1.5$, which is evident also at lower r_{sky} . This minimum is caused by WBs having a sharp peak to their \tilde{v} distribution at $\tilde{v} \approx 0.5$ followed by a rapid decline as a consequence of orbital mechanics and projection effects (e.g. BZ18, PS19). There is an extended tail due to CB contamination, but this too

would generally lead to a declining \tilde{v} distribution because a slower CB orbital velocity corresponds to a wider CB separation, which is less likely (a more detailed explanation is given in Banik, Pittordis & Sutherland 2021). However, we expect an increasing contribution from chance alignments at larger \tilde{v} because their distribution is expected to be linear in \tilde{v} [see equation (27)]. The minimum in the \tilde{v} distribution is not very apparent at $r_{\text{sky}} \lesssim 5$ kAU, where WBs are very common and chance alignments are relatively unimportant. The minimum becomes readily apparent at larger r_{sky} where WBs are quite rare, increasing the relative contribution of chance alignments because they should have a flat distribution with respect to r_{sky} [equation (27)]. The fact that WBs are much less common at high r_{sky} naturally leads to empty pixels around the minimum of the \tilde{v} distribution, with the \tilde{v} range covered by these empty pixels expected to gradually widen with increasing r_{sky} .

It is clear that we need a detailed model including CB and LOS contamination if we are to infer the law of gravity at low accelerations and search for the predicted MOND signal. This is the subject of the next section.

3 MODELLING THE WB DATA SET

The $(r_{\text{sky}}, \tilde{v})$ distribution of our WB sample is a consequence of several simultaneously occurring physical processes and projection effects, which we try to model in this section to best reproduce Fig. 6. In addition to WB orbital motion (Section 3.1), we also consider how an undetected CB population affects the observables (Section 3.2) and take into account the possibility of some LOS contamination (Section 3.3). This might be field stars, but it could also be from stars that formed in the same star cluster. Given the results of PS19 and the excellent quality of *Gaia* DR3, we expect LOS contamination to only slightly affect our WB sample and to become important only at large separation and \tilde{v} .

LOS contamination can be included in our model fairly easily. Moreover, detailed WB orbit modelling has previously been conducted in Newtonian and Milgromian gravity in a manner that can readily be repurposed for the WBT (BZ18). As a result, the main complication that we need to handle is the possibility of an undetected CB companion around 0, 1, or 2 of the stars in a WB. Since recoil motion and additional mass created by an undetected CB physically affect the WB rather than merely contribute an extra population to the statistics, including CB contamination greatly increases the complexity and computational cost of the WBT. We were eventually able to devise a detailed plan that captures the essential physics in a computationally feasible way (Banik, Pittordis & Sutherland 2021). We refer the reader to that work for a more detailed look at the computational techniques needed to implement the WBT in a reasonable timeframe. This allows us to prepare the first detailed model of the WB population whose parameter space is thoroughly explored using standard statistical techniques.

3.1 The WB population

We set up a library of WB orbits using the prior work of BZ18 with minimal adjustments, which we briefly describe. We need to run calculations over a wider mass range in order to cover the range of WB masses in our sample (red bars in Fig. 5). We also found that the computational cost of including CB contamination can be greatly reduced if instead of recording results with respect to the projected separation r_{sky} , we use

$$\tilde{r}_{\text{sky}} \equiv \frac{r_{\text{sky}}}{a}, \quad (11)$$

where a is the semi-major axis of the WB (this is motivated in Section 3.2.3). An incidental benefit of this approach is that at low a , recording probabilities into different bins in \tilde{r}_{sky} rather than r_{sky} improves the resolution in r_{sky} . Since we record results separately for different a and only marginalize over a at a much later stage, we can recover $r_{\text{sky}} \equiv a\tilde{r}_{\text{sky}}$ as and when needed, which is necessary for a comparison with observations because a is generally not known.

In MOND, the Galactic EFE plays a crucial role (BZ18). This can be estimated as $g_e = v_{c,\odot}^2/R_0$, where R_0 is the Galactocentric distance of the Sun and $v_{c,\odot}$ is the Galactic rotation curve amplitude at the Solar circle (sometimes called the Local Standard of Rest or LSR). As in section 3.6 of BZ18, we assume that $v_{c,\odot} = 232.8 \text{ km s}^{-1}$ and $R_0 = 8.2 \text{ kpc}$ (McMillan 2017), which implies that $g_e = 1.785 a_0$. Subsequent studies show that R_0 must be very close to 8.2 kpc based on combining astrometry and spectroscopy of the star S2 near the Galactic centre black hole (Gravity Collaboration 2019). Since its proper motion is precisely known (Reid & Brunthaler 2004; Gordon, de Witt & Jacobs 2023), the Galactocentric tangential velocity of the Sun now has very little uncertainty, constraining the LSR velocity to very close to 233 km s^{-1} (McGaugh 2018; Zhou et al. 2023) if the non-circular motion of the Sun is taken to be precisely known from Schönrich, Binney & Dehnen (2010). The latter is somewhat more uncertain than the other parameters entering into the LSR velocity (Schönrich 2012; Francis & Anderson 2014; Bovy et al. 2015), but it is still clear that $v_{c,\odot}$ is very unlikely to differ from our assumed value by more than 10 km s^{-1} , which implies that uncertainty in $v_{c,\odot}$ has little effect on the predicted MOND signal in local WBs (Section 5.3). Moreover, g_e has recently been directly determined by the measurement of the Solar System’s acceleration relative to distant quasars (Gaia Collaboration 2021b). Their work discovered the changing aberration angle of distant quasars due to the changing velocity of the Solar System barycentre. Its directly measured acceleration points directly towards the Galactic centre within a small uncertainty (see their fig. 10) and has a magnitude of $g_e = 1.94 \pm 0.13 a_0$. Since gravitational fields from objects beyond the Galaxy are thought to be negligible in comparison to the Galactic gravitational field, this provides a completely independent and very direct confirmation of the kinematic estimate of g_e used in BZ18 and in this contribution.

3.1.1 Orbit modelling

WBs are bound together by their mutual gravity. This is easy to calculate in the Newtonian case but is somewhat more complicated in MOND, as discussed next. Our calculation of the Milgromian gravitational field and integration of WB orbits in this field use the same approach as BZ18, to which we refer the reader for further details.

In an isolated spherically symmetric mass distribution, the asymptotic MOND behaviour is given by equation (1). To interpolate between the Newtonian and deep-MOND regimes, we have to use an interpolating function ν with argument g_N such that $\mathbf{g} = \nu \mathbf{g}_N$. In the quasilinear formulation of MOND (QUMOND; Milgrom 2010), this can be generalized to a more complicated geometry by taking the divergence of both sides.

$$\nabla \cdot \mathbf{g} = \nabla \cdot (\nu \mathbf{g}_N). \quad (12)$$

In this contribution, we adopt the simple interpolating function (Famaey & Binney 2005) for reasons discussed in section 7.1 of

BZ18.

$$\nu = \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{a_0}{g_N}}. \quad (13)$$

The WBT depends significantly on the adopted interpolating function because local WBs have an intermediate total acceleration. Using a very sharp transition can cause WBs to be completely Newtonian even if MOND is correct. However, this would not be consistent with the RAR (we discuss this further in Section 5.3).

Since local WBs are significantly affected by the Galactic EFE, we need to include the Galactic g_e . This is assumed to point directly towards the Galactic centre and to have a magnitude of $1.785 a_0$ based on the Galactic rotation curve, as discussed above. In QUMOND, we need the Newtonian-equivalent external field $\mathbf{g}_{N,e}$, which is what the Galactic gravity on the Solar neighbourhood would have been in Newtonian gravity without DM. As discussed in section 9.3.1 of BZ18, we can use the spherically symmetric relation between \mathbf{g}_e and $\mathbf{g}_{N,e}$ because the Sun is many disc scale lengths from the Galactic centre, reducing the importance of the Galaxy’s disc geometry.

$$g_e \nu(g_e) = \mathbf{g}_{N,e}. \quad (14)$$

Solving this implicitly gives $g_{N,e} = 1.144 a_0$ in the Solar neighbourhood. The Newtonian gravity in a WB is then found by adding $\mathbf{g}_{N,e}$ to the contribution from the WB itself, which is called $\mathbf{g}_{N,i}$ in some works as it is internal to the WB.

To limit the computational cost, we treat each WB as a test particle orbiting a point mass containing the binary total mass M . This approach is valid in Newtonian gravity, but it was shown in section 7.3 of BZ18 that it remains a very good approximation in MOND. This is because the fairly strong Galactic EFE causes local WBs to lack an extended regime in which they are isolated and at low accelerations. For all intents and purposes, WBs are either in the Newtonian regime or they are dominated by the EFE. In either case, MOND becomes linear in the mass distribution, allowing the potentials of the two stars to be superposed.

With this approximation, the problem becomes axisymmetric about g_e . We then solve equation (12) using the ‘ring library’ protocol discussed in BZ18. This involves finding the source term $\nabla \cdot (\nu \mathbf{g}_N)$ at every point in a 2D grid and then using direct summation to find \mathbf{g} . The calculations are accelerated at off-axis locations by scaling the previously calculated Newtonian gravity of a unit mass and unit radius ring, avoiding the need to sum contributions to \mathbf{g} from different positions around each ring. An analytic allowance is made for $\nabla \cdot (\nu \mathbf{g}_N)$ being non-zero outside the rather large region used in the integration (the extra contribution to the potential is given in equation 9 of BZ18; see their appendix A for a derivation).

The Milgromian potential of a point mass is not spherically symmetric in the presence of the EFE (Milgrom 1986a, 2010; Banik & Zhao 2018a). This has some interesting consequences (Candlish et al. 2018; Thomas et al. 2018; Banik & Kroupa 2019b; Banik et al. 2020; Kroupa et al. 2022). Since the WB separation r is expected to sample a range of directions relative to the EFE, the WBT is mostly sensitive to the angle-averaged inward radial gravity, which we denote $\langle g_r \rangle$. Since the use of \tilde{v} rather than actual velocities effectively divides out the Newtonian prediction g_N , the WBT is really about the parameter

$$\eta \equiv \frac{\langle g_r \rangle}{g_N}, \quad (15)$$

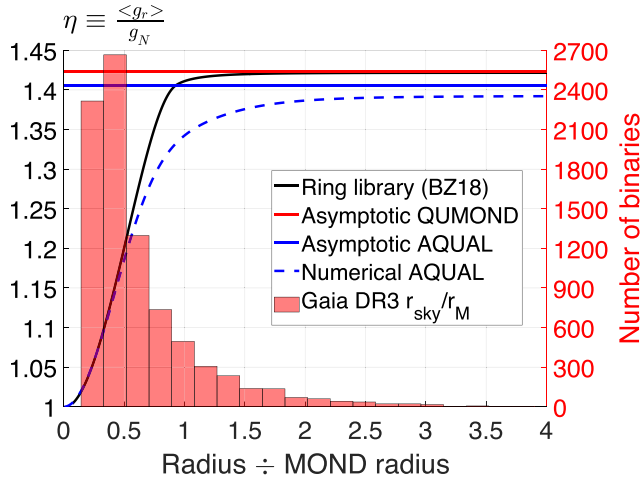


Figure 7. The azimuthally averaged boost factor η to the radially inward gravity compared to the Newtonian expectation for WBs with different separation relative to their MOND radius, assuming the simple interpolating function [equation (13)]. The solid red (blue) horizontal line shows the asymptotic analytic expectation in QUMOND (AQUAL), which we give in equation (37) [equation (38)]. The numerical QUMOND result central to our study (solid black line; BZ18) has the expected asymptotic behaviour. The dashed blue line shows equation 15 of Chae & Milgrom (2022), which has the wrong asymptotic limit (mismatch between blue lines towards the right) and is based on numerical calculations that only deal with a weak external field rather than a dominant one (see the text). The red histogram shows the distribution of r_{sky}/r_M for the WBs in our sample. The outer edges of the most extreme bins show the full range of the data set.

whose square root is probed by velocity measurements. We use Fig. 7 to show η as a function of r/r_M (solid black line).⁶ There is no boost to the Newtonian expectation when $r \ll r_M$, but this changes rapidly until $r \approx r_M$, beyond which the Milgromian enhancement to $\langle g_r \rangle$ saturates at just over 40 per cent due to the EFE. Without it, the boost factor would continue growing linearly with r . In the asymptotic or quasi-Newtonian regime (discussed further in section 3.3 of Asencio et al. 2022) where the enhancement factor has reached its maximum, the WB gravity can be treated as a small perturbation about the Galactic gravity, making the potential analytic (equation 30 of Banik et al. 2020). The analytic expectation is shown using a horizontal red line, revealing an excellent match to the numerical result when $r > r_M$.

In this contribution, we focus on the QUMOND formulation because it is more computer-friendly than the earlier quadratic Lagrangian formulation (AQUAL; Bekenstein & Milgrom 1984). The asymptotic AQUAL result is shown in Fig. 7 using a horizontal blue line (equation 35 of BZ18). As discussed there, this result is very similar to the QUMOND result for the same interpolating function, i.e. the same relation between \mathbf{g} and \mathbf{g}_N in spherical symmetry. Numerical AQUAL results are not available in the transition zone where the EFE is neither negligible nor dominant, but we can get a general idea using equation 15 of Chae & Milgrom (2022), which is a fit to numerical results for a weak EFE as relevant to galaxy rotation curves. We show this using a dashed blue line. It is clear that this equation has the wrong asymptotic limit, as can also be seen from the equation: when the EFE dominates, it takes the argument $1.1 g_{N,e}$ rather than $g_{N,e}$. Even if this issue is fixed, the functional

⁶Zonoozi et al. (2021) provide an analytic fit to these results for arbitrary g_e with the correct asymptotic behaviour (see their equations 23 and 24).

form is not compatible with the EFE-dominated point mass potential in AQUAL (equation 66 of Milgrom 2010) or the resulting radial gravity (equation 32 of BZ18). Despite these issues when the EFE is strong, equation 15 of Chae & Milgrom (2022) should still give a reasonable idea of the AQUAL enhancement to $\langle g_r \rangle$ when $r \ll r_M$ because it is a fit to numerical results in this regime. We see that when $r \lesssim 0.5 r_M$, the AQUAL and QUMOND results are almost identical. Therefore, results for the WBT are not sensitive to which formulation of MOND we use. We discuss this further in Section 5.3.

Another aspect of Fig. 7 is the distribution of r_{sky}/r_M , which we show using red bars (the left edge of the first bar shows the minimum value). Our sample covers down to $r_{\text{sky}}/r_M = 0.14$, so we consider it to be reasonably sensitive down to $r/r_M = 0.2$, leaving some allowance for 3D separations exceeding r_{sky} . A WB separated by only $r_M/5$ has an acceleration of $g_N = 25 a_0$ and can thus be considered isolated. In this case, $v = 1/2 + \sqrt{1/4 + 1/25} = 1.04$, which is very small compared to the predicted enhancement to g_N for more widely separated WBs. This is even truer for interpolating functions that have a somewhat faster return to Newtonian behaviour at high accelerations in order to better match Solar System ephemerides (Hees et al. 2014, 2016). At the other end, our sample probes well into the asymptotic regime where we would expect almost the maximum possible enhancement to g_N given the Galactic EFE. Our sample thus covers a wide enough range of internal WB accelerations to allow a sensitive search for the predicted MOND effect. While our WBs experience a narrower range of total gravitational field strengths because the Galactic EFE essentially imposes a lower limit, we argue later that this is not a major limitation given rotation curve constraints on the MOND interpolating function in the transition zone (Section 5.3).

We use the assumed gravity law to integrate WB orbits with a range of mass M , semi-major axis a , and orbital eccentricity e , defining a and e for a generalized gravity theory as in section 2.3.1 of BZ18. Following their approach, each orbit is started at a separation of a and integrated for 20 revolutions. This should be more than sufficient in the Newtonian case, but it is also adequate in MOND because the resulting statistical distribution is similar to that yielded by a much longer 5 Gyr integration (see their fig. 15). We also consider a range of inclinations i between the orbital pole and the Galactic centre direction defining the EFE. Results for different inclinations are marginalized assuming a prior of $\sin i$, which is appropriate for an isotropic distribution of orbital poles. Orbits are truncated if the stars get to within 50 AU or become more distant than 100 kAU as we consider such WBs to have crashed together or to have likely become unbound at some point by perturbations from passing stars, respectively (section 2.3.2 of BZ18). At each time step, we consider a dense 2D grid of viewing directions and compute the $(\tilde{r}_{\text{sky}}, \tilde{v})$ inferred by a distant observer along the chosen LOS, with results statistically weighted by the solid angle covered by each LOS. We also weight the results at different time steps by the duration of the time step, which varies because we use an adaptive time step to reduce the computational cost. In this way, we build up a simulated $(\tilde{r}_{\text{sky}}, \tilde{v})$ distribution for different (M, a, e) .

In the Newtonian case, it is not necessary to consider different M or i , but we use a $5 \times$ higher resolution in e because the \tilde{v} distribution at fixed e has sharp peaks corresponding to pericentre and apocentre and e does not change in a Newtonian orbit integration (unlike in MOND; see Paučo & Klačka 2016). To get a smooth distribution when marginalizing over different e , BZ18 found that a rather high resolution is required in e for the Newtonian case. This is not needed in MOND because the peaks are blurred out by changes in the shape of the orbit over time: typically about 8 orbits are needed to go from

nearly circular to very eccentric back to nearly circular (see their fig. 20).

3.1.2 Orbital parameter distribution

The mass distribution of WBs is obtained directly from that of our WB sample (Fig. 5). We count up how many WBs are in the mass range covered by each value of M at which we perform an orbit integration. This directly determines the relative statistical weights of different M values in our WB orbit library.

Following Andrews, Chanamé & Agüeros (2017), we assume that the semi-major axes of the WBs follow a broken power law in a . The logarithmic slope at low a is fixed to -1.6 , which is known to be valid down to at least 0.5 kAU (El-Badry & Rix 2018). While this can in principle extend up to arbitrarily large a , we assume that it breaks down at $a = a_{\text{break}}$, beyond which the logarithmic slope becomes $\beta < -1$ to ensure a convergent integral. Our assumed distribution of a is therefore given by

$$P(a) \propto \begin{cases} a^{-1.6}, & \text{if } a < a_{\text{break}}, \\ a^{\beta}, & \text{if } a > a_{\text{break}}, \end{cases} \quad (16)$$

where a_{break} and β are free parameters that our analysis infers from the data. Note that since we are modelling the distribution of our sample of WBs rather than the full WB population, the inferred values of these parameters may be somewhat biased. In particular, the requirement to have accurate \tilde{v} measurements is quite difficult to satisfy at large separations because the Newtonian v_c is low. This is likely to reduce β .

Following Hwang, Ting & Zakamska (2022), we parametrize the distribution of orbital eccentricities using a power law with index γ .

$$P(e) = (\gamma + 1)e^{\gamma}. \quad (17)$$

γ is left as a free parameter that is allowed to vary over the range 0 – 4 in all our analyses. We expect a nearly thermal distribution of eccentricities ($\gamma = 1$; Jeans 1919; Ambartsumian 1937) – this is discussed in more detail in Section 4.2 of Kroupa (2008). We use a flat prior on γ for our nominal analysis. However, Hwang, Ting & Zakamska (2022) found that a slightly superthermal distribution ($\gamma > 1$) better fits the observed distribution of the angle ψ between r_{sky} and v_{sky} (this is related to the WB orthogonality test outlined in section 8.2.1 of BZ18). We therefore run a variation of our nominal analysis in which we assume a Gaussian prior of $\gamma = 1.32 \pm 0.09$, which should shrink the uncertainties somewhat. We note that since CB contamination would tend to flatten out the distribution of ψ and thereby drive the inferred eccentricity distribution towards thermal, it is quite likely that γ is slightly higher than inferred by Hwang, Ting & Zakamska (2022) given that those authors do not consider undetected CBs. On the other hand, their fig. 10 shows that γ can be overestimated slightly at large separations for an underlying slightly superthermal distribution.

3.1.3 Interpolating the gravity law

An important aspect of our analysis is that we consider an arbitrary gravity law on a sliding scale between Newtonian and Milgromian. For this purpose, we introduce the gravity law parameter α_{grav} , which is the most important parameter for the WBT. We define this such that $\alpha_{\text{grav}} = 0$ represents Newtonian gravity and $\alpha_{\text{grav}} = 1$ represents QUMOND with the simple interpolating function. We allow α_{grav} values somewhat outside this range to capture the possibility of behaviour different to either theory. The most plausible outcomes

to the WBT are a strong preference for α_{grav} values very close to 0 or 1, with the large sample size hopefully tightening the error budget enough that both possibilities are not simultaneously consistent with the observations (a few thousand WBs should be sufficient; see BZ18). Since the WBT will have some uncertainties, we allow α_{grav} to lie in the range $(-2, 3.6)$. Negative values imply that the gravity in the asymptotic regime has a lower than Newtonian normalization, while $\alpha_{\text{grav}} > 1$ implies a normalization higher than expected in MOND.

To set up a parametrized gravity law, we obtain the simulated $(\tilde{r}_{\text{sky}}, \tilde{v})$ distribution in both gravity theories for each (M, a, γ) , with different orbital eccentricities marginalized over for the adopted γ using the eccentricity distribution it defines [equation (17)]. The main idea of our parametrization is to transform the Newtonian $(\tilde{r}_{\text{sky}}, \tilde{v})$ distribution into the Milgromian one and then interpolate the coefficients involved in the transformation so that arbitrary gravity laws can be considered.⁷ To simplify our analysis, we consider a simple 2D stretch in which the probability distribution is stretched along the \tilde{r}_{sky} axis by a factor S_r and then along the \tilde{v} axis by a factor S_v , with $S_r \neq S_v$ in general. These are found by running a 2D gradient descent in (S_r, S_v) to minimize the sum of squared residuals between the stretched Newtonian array and the MOND array (Fletcher & Powell 1963). While a simple 2D stretch applied to the former does provide a good approximation to the latter, there are still some residuals, which we calculate for each pixel. This array of residuals is itself scaled by a factor of α_{grav} and stretched in 2D by some factors depending on α_{grav} and (S_r, S_v) . The idea behind applying this ‘corrections array’ is to recover the Newtonian (Milgromian) $(\tilde{r}_{\text{sky}}, \tilde{v})$ distribution exactly if $\alpha_{\text{grav}} = 0$ (1). If we consider an intermediate case where $\alpha_{\text{grav}} = 0.5$, we would expect the \tilde{v} distribution to stop at a value about halfway between the maximum which arises in the Newtonian and MOND distributions. To prevent the corrections array from creating non-zero contributions all the way out to the maximum \tilde{v} which arises in MOND, the residuals array itself needs to be stretched somewhat, though in this case we expect $S_v < 1$. Since no corrections are required if $\alpha_{\text{grav}} = 0$, we need to avoid applying any corrections in this case, which is achieved by our scaling factor of α_{grav} . We use linear interpolation in all the stretch and scaling factors discussed above with respect to α_{grav} . To avoid negative probabilities due to negative values in the corrections array, any negative entries in the resulting $(\tilde{r}_{\text{sky}}, \tilde{v})$ distribution are converted to zero and the array renormalized.

3.2 Undetected companions

As discussed in Section 1.1, undetected companions are the main source of contamination to the WBT given the low measurement uncertainties (Fig. 4) and the low LOS contamination fraction evident from the low number counts in the rather wide pixels at high r_{sky} and \tilde{v} (Fig. 6). In particular, the declining \tilde{v} distribution it reveals is a clear sign of some process that broadens the WB distribution, which we otherwise expect to be concentrated at $\tilde{v} \lesssim 2$ for any plausible gravity theory (PS19).

We therefore need to include the possibility that a star in a WB is actually itself part of a CB, even though the CB companion is undetected. Including CBs proved to be by far the most challenging aspect of the WBT, so we had to make some simplifying assumptions to keep the complexity and computational cost manageable. Due to their much smaller separation, we assume that the CB is completely

⁷We marginalize over M and a at a later stage.

Newtonian internally, even if MOND affects the motion of its barycentre around the other star in the WB.⁸ It will become clear later that the undetected CBs relevant to the WBT indeed have only a small separation compared to the WB, justifying this approximation.

3.2.1 Impact on WB observables

Undetected companions have two main effects on a WB: increased mass and induced recoil velocity. Since the CB companion is expected to tightly orbit the star it contaminates, we neglect the impact of the CB on the WB's r_{sky} . For this reason, we also assume that the light emitted by the 'undetected' star is blended with that of the contaminated star, i.e. they are unresolved by *Gaia*. This is reasonable given that resolved third star companions would cause the WB to be removed from our sample at an early stage (Section 2.1). Indeed, we will see later that our analysis prefers to avoid CBs whose separation is significant compared to that of the WB.

The actual mass of the CB exceeds its estimated mass by a factor

$$\Delta\tilde{M} \equiv \frac{\text{Total CB mass}}{\text{Estimated CB mass}} - 1. \quad (18)$$

This is positive because the mass–luminosity relation is very steep (Fig. 1), so blended light from the 'undetected' companion raises the estimated CB mass by less than the companion's mass. As a result, two unresolved Sun-like stars would appear to have a total mass of only $1.19 M_{\odot}$ even though their total mass is actually $2 M_{\odot}$, yielding $\Delta\tilde{M} = 0.68$. Since the mass–luminosity relation is not an exact power law, $\Delta\tilde{M}$ depends on the masses of the stars involved. We handle this complication by considering as a prior the mass distribution of all the stars in our WB sample (blue bars in Fig. 5). For a fixed fraction $\tilde{q} \leq 1/2$ of the CB total mass in the undetected companion, we consider a dense grid of primary star masses covering the range shown there.⁹ In each case, we use the mass–luminosity relation from Pecaut & Mamajek (2013) to obtain the fraction \tilde{L} of the CB's total luminosity contributed by the undetected companion – this will be important later.¹⁰ We then numerically invert equation (6) using the Newton–Raphson algorithm to find the estimated mass of the CB given the total *Gaia*-band luminosity of the two stars in it, thus following the same approach as observers would use without knowing about the CB. In this way, we obtain the relation between $\Delta\tilde{M}$ and \tilde{q} , which we show as the solid red line in Fig. 8. We also use a dotted red line to show the results if we neglect the blended light contributed by the undetected companion. The results are similar either way unless the CB consists of two stars with an almost equal mass: in this limit, blended light reduces $\Delta\tilde{M}$ from 1 to 0.72.

The impact of blended light is more significant when it comes to the recoil velocity induced by an undetected companion. In the limit that the CB has two equal mass stars such that $\tilde{q} = 1/2$, the photocentre and barycentre would coincide and move only due to the WB orbit. This is true regardless of the relative velocity \mathbf{v} of the CB orbit. In general, the velocity of the primary star relative to the CB barycentre would be $\tilde{q}\mathbf{v}$, while the undetected star would move at $-(1-\tilde{q})\mathbf{v}$, which is in the opposite direction. Once these velocities are averaged weighted by the mass fractions, the CB barycentre is of course static.

⁸Bekenstein & Milgrom (1984) showed that a body with high internal accelerations can still exhibit MONDian behaviour for its barycentre.

⁹'Primary' here refers to the more massive star in the CB, which may itself be a sub-dominant component of the total WB mass.

¹⁰For simplicity, we do not implement the FLAME calibration discussed in Section 2.3.3 when converting masses to luminosities or vice versa in this section.

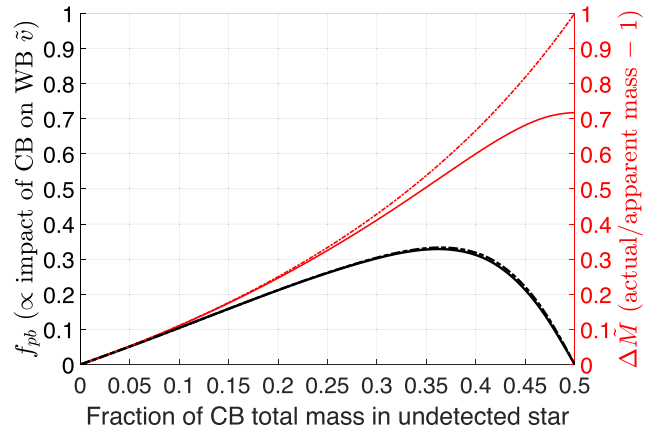


Figure 8. How the fraction of a CB's total mass in its undetected component affects the offset between its photocentre and barycentre (black lines) and the fraction by which the total mass exceeds the mass estimated by observers based on the luminosity (red lines). The scale for the left y-axis is arbitrary: the important aspect is the functional form. The dot–dashed lines neglect the increase to the estimated mass caused by the luminosity of the undetected component. The solid lines include this effect, which helps to somewhat mitigate the impact of CB contamination. The solid black line is proportional to the CB's impact on the WB \tilde{v} under the assumption that an undetected CB increases the estimated mass of both WB components by a similar fraction. Only the solid lines shown here enter into our model as $\Delta\tilde{M}$ [red; equation (18)] and f_{pb} [black; equation (20)].

But if we instead weight the stars using their luminosities, we get that the photocentre velocity is

$$\mathbf{v}_{\text{phot}} = \left[\tilde{q} (1 - \tilde{L}) - \tilde{L} (1 - \tilde{q}) \right] \mathbf{v} = (\tilde{q} - \tilde{L}) \mathbf{v}. \quad (19)$$

Since we are only interested in the magnitude of \mathbf{v}_{phot} , the impact on the observed WB relative velocity (based on photocentres) is $\propto |\tilde{q} - \tilde{L}|$ (as also found by Penoyre, Belokurov & Evans 2022). Moreover, \mathbf{v} is itself proportional to the Keplerian velocity of the CB, which depends on its total mass and separation. Finally, the impact on \tilde{v} also needs to take into account that blended light increases the apparent mass of the WB and thus its Newtonian v_c [equation (3)], but the extra mass is underestimated thanks to the steep mass–luminosity relation. We assume for simplicity that both stars in a WB are affected by CBs with a similar mass ratio to their primary star.

Putting all this together, we get that the impact on \tilde{v} scales with the parameter

$$f_{\text{pb}} = \left| \tilde{q} - \tilde{L} \right| \sqrt{1 + \Delta\tilde{M}}. \quad (20)$$

This is shown as the solid black line in Fig. 8. The dotted black line shows the result if $\Delta\tilde{M}$ is calculated neglecting blended light. It is clear that f_{pb} is not sensitive to details of how $\Delta\tilde{M}$ is calculated for the purposes of the above equation. This is because if $\tilde{q} \ll 1/2$ and the CB has an extreme mass ratio, then the undetected star contributes almost no blended light, making $\Delta\tilde{M}$ insensitive to whether we consider this light. In the opposite limit that $\tilde{q} \approx 1/2$, the stars in the CB are very similar such that $\tilde{q} \approx \tilde{L}$, making f_{pb} very small and thus rendering irrelevant small differences in f_M .

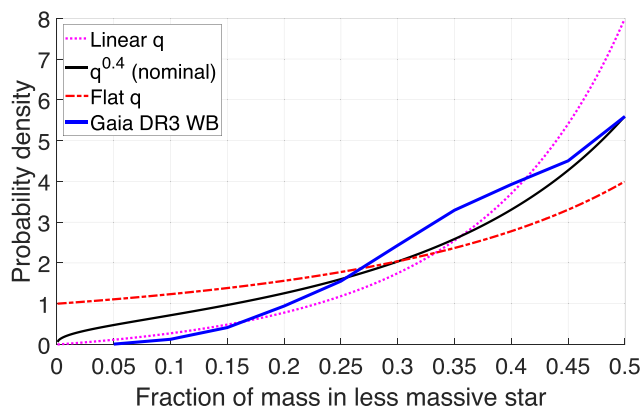


Figure 9. The mass ratio distribution of the WBs in our *Gaia* DR3 sample (thick solid blue line). We consider three different forms for the undetected CB population. The nominal assumption of a $q^{0.4}$ dependency (solid black line) is based on Korntreff, Kaczmarek & Pfalzner (2012), with $q \leq 1$ being the mass ratio between the stars in each binary. This model fits the WB distribution fairly well, especially given that selection effects would make it hard to find binaries with $q \ll 1$. We also consider the case of a flat distribution in q (dot-dashed red line) and a linear distribution (dotted magenta line). These should bracket the range of possible q distributions.

3.2.2 Mass ratio distribution

The impact of an undetected CB on the WB observables depends on the fraction of the CB mass in the undetected star (Fig. 8). This requires us to assume some distribution for $\tilde{q} \leq 1/2$, or equivalently for the mass ratio $q \equiv \tilde{q}/(1 - \tilde{q}) \leq 1$. Following Korntreff, Kaczmarek & Pfalzner (2012), we assume that

$$P(q) \propto q^{0.4}, \quad q \leq 1. \quad (21)$$

We expect this to broadly match the distribution of WB mass ratios. We therefore use Fig. 9 to show the mass fraction in the less massive star, which is the quantity more relevant for our analysis [equation (24)]. The $q^{0.4}$ assumption is shown as a solid black line, while the distribution for our WB sample is shown as a solid blue line. It is apparent that our adopted $q^{0.4}$ distribution provides a good match to the WB distribution. The WB sample contains a deficit of systems with a very extreme mass ratio, but this is almost certainly due to selection effects: low mass stars are very faint (Fig. 1). Even if they are not present in our WB sample, we still need to consider extreme mass ratio CBs in our analysis because they can affect the WB observables if the CB has a small separation.

To bracket the possible uncertainties in the CB mass ratio distribution, we also consider a flat distribution of q and a linear distribution, which we show on Fig. 9 as the dot-dashed red and dotted magenta line, respectively. A flat distribution is a rough approximation to the three-part power law recommended in fig. 7 of El-Badry et al. (2019). We will see later that using instead a flat or linear prior on q gives similar results to our nominal $q^{0.4}$ assumption, so our conclusions are similar for any power-law distribution of q with exponent in the range 0–1 (Section 5.1.2).

3.2.3 Semi-major axis distribution

The Keplerian orbital velocity of the CB is affected by its semi-major axis a_{int} . Following PS23, we impose an upper limit of ak_{CB} , where a is the semi-major axis of the WB. Thus, we require that

$$\tilde{a}_{\text{int}} \equiv \frac{a_{\text{int}}}{a} < k_{\text{CB}}. \quad (22)$$

We work with \tilde{a}_{int} rather than a_{int} because the ratio between the CB and WB orbital velocities depends on the ratio of their semi-major axes. The maximum allowed value of this ratio is k_{CB} , which is a free parameter in our analysis.

Since we remove systems with an obvious third star, k_{CB} refers to the maximum \tilde{a}_{int} for CB companions to the WBs in our sample, bearing in mind that we do not consider stars which have a discernible companion within 0.5 arcsec with a reasonably consistent parallax and proper motion (Section 2.1). To convert this to a separation in AU, we note that as our sample goes out to 250 pc, we can expect a reasonable number of systems out to 200 pc. Thus, we expect WBs in our sample to typically not have CB companions with projected separation beyond 100 AU. The upper limit to the actual separation would be larger thanks to projection effects, but their impact on a_{int} would be counteracted by the high likelihood of the orbit being quite eccentric (Hwang, Ting & Zakamska 2022) and the fact that orbital phases near apocentre are more frequent. It is possible for a sufficiently faint star to go undetected at an even larger separation, but it would have to evade our faint star catalogue used to reduce CB contamination (Section 2.1). This catalogue goes down to an apparent magnitude of $m_G = 20$, which at a distance of 200 pc corresponds to $M_G = 20 - 5 \log_{10} 20 = 13.5$ and thus a mass of about $0.1 M_{\odot}$ (Fig. 1). As a result, CB companions beyond 0.5 arcsec from the contaminated star can remain undetected only in a narrow range of mass at the very bottom of the main sequence. Such a low mass would also reduce the perturbation to the velocity of the contaminated star [low f_{pb} in equation (20)]. Combined with the large separation, this makes it unlikely that a distant low-mass companion which evades our faint star catalogue would actually be relevant to our analysis. We thus assume that the maximum relevant $a_{\text{int}} \approx 100$ AU. To convert this to k_{CB} , we note that a reasonable a for the WBs might be 40 kAU because our WB sample goes up to $r_{\text{sky}} = 30$ kAU and we need to consider all viewing angles and orbital phases. For this reason, our WB orbit library goes up to $a = 57$ kAU, though we expect our analysis to assign only very small statistical weights at such high a due to its declining distribution [equation (16)] and the fact that we truncate orbits when they go beyond 100 kAU to mimic disruption caused by Galactic tides and encounters with other stars. Taking the ratio of the above estimates for a_{int} and a suggests that $k_{\text{CB}} \approx 2.5\%$. We do not impose this in our analysis, but we do set a prior that it should be at least 1‰ to avoid numerical difficulties. This is very conservative because it implies a maximum a_{int} of about 40 AU, which corresponds to just 0.2 arcsec for a WB 200 pc away. One can easily envisage that a CB with a larger separation would not be detected at this distance.

At the opposite end, the lowest a_{int} that we need to consider is set by the fact that CBs on very small orbits have a short orbital period, leading to rapid but low amplitude astrometric oscillations that have little impact on the inferred space velocity of the contaminated star. This phenomenon is not directly included in our analysis, which for simplicity assumes that the stars in a CB move at constant velocity over the whole *Gaia* observing baseline. In Appendix B, we estimate that this approximation breaks down when $a_{\text{int}} \lesssim 3.2$ AU, below which the impact of the CB rapidly becomes much smaller than our estimate based on uniform linear motion.

Based on the above considerations, we allow CBs to have a 1.5 dex range of \tilde{a}_{int} with a maximum at $k_{\text{CB}} > 1\%$. We assume that \tilde{a}_{int} has a flat distribution in log-space, as in the classical Öpik law (Öpik 1924). The logarithmic distribution of a_{int} does indeed seem to be rather flat between 10 and 100 AU (fig. 2 of Offner et al. 2023), so this ought to be a reasonable assumption. We will see later that our

results are not much affected by using a somewhat different shape or width to the a_{int} distribution (Section 5.1.3).

In addition to properties of the CB, how it affects \tilde{v} of the WB also depends on the latter's projected separation. At fixed a , the impact on \tilde{v} is reduced at low r_{sky} because this raises the predicted Newtonian v_c of the WB [equation (3)]. Considering also that the CB orbital velocity $\propto 1/\sqrt{a_{\text{int}}}$, the perturbation to the WB \tilde{v} scales as

$$\Delta\tilde{v} \propto \sqrt{\frac{\tilde{r}_{\text{sky}}}{\tilde{a}_{\text{int}}}}. \quad (23)$$

The lack of any explicit dependence on a greatly reduces the computational cost of the WBT, but this is only possible because the WB libraries are stored with respect to \tilde{r}_{sky} rather than r_{sky} . The only slight downside is that we need the latter for a comparison with observations, so we subsequently need to find $r_{\text{sky}} \equiv a\tilde{r}_{\text{sky}}$.

3.2.4 Convolution of the CB and WB libraries

The fractional increase to the total mass of a WB is found by combining $\Delta\tilde{M}$ for each of the stars in it, which we denote using A and B subscripts.

$$\Delta\tilde{M} = \tilde{q}_A \Delta\tilde{M}_A + \tilde{q}_B \Delta\tilde{M}_B, \quad (24)$$

where $\tilde{q}_A \equiv 1 - \tilde{q}_B$ is the fraction of the WB total mass in star A and $\Delta\tilde{M}_A$ is the fraction by which its actual mass exceeds its apparent mass [equation (18)]. If one of the WB stars is not in a CB, then $\Delta\tilde{M} = 0$ for that star. We assume that \tilde{q}_A and \tilde{q}_B have an identical distribution, which we obtain from our WB sample (solid blue line in Fig. 9). This is not exactly correct because undetected CBs can alter the mass ratio. However, we expect this to be only a very minor issue because even in the extreme case of having two equal Solar mass stars in a WB where one of the stars has an undetected equal mass companion, observers would infer a mass fraction in the less luminous component of 0.46 rather than the correct value of 1/3. The error would be smaller in more typical cases, especially as both stars in the WB could actually be CBs. We therefore expect the WB mass ratio inferred from luminosities to be quite accurate.

To obtain the relative velocity between the photocentres and barycentres of the CBs making up a WB, we again need to combine the contributions from both CBs. In this case, the scaling factors are $\sqrt{\tilde{q}_A}$ and $\sqrt{\tilde{q}_B}$ because the Keplerian orbital velocity of a CB scales with the square root of its mass. Allowing also for the increased WB orbital velocity due to its mass being higher than estimated, we get that the WB \tilde{v} should be revised as follows:

$$\tilde{v} \rightarrow \tilde{v} f_M + \Delta\tilde{v}_A \sqrt{\tilde{q}_A} + \Delta\tilde{v}_B \sqrt{\tilde{q}_B}, \quad (25)$$

where f_M was defined in equation (20) and $\Delta\tilde{v}_A$ is the change to \tilde{v} of the WB if all its mass was in star A such that $\tilde{q}_A = 1$ ($\Delta\tilde{v}_B$ is defined analogously). Since $\Delta\tilde{v}$ induced by each CB is a 2D vector, we also need to allow for partial cancellation between the photocentre-barycentre offsets arising from each CB. This leads to 4 possible outcomes for whether the contributions along each direction are parallel or opposite.

Following El-Badry et al. (2019), we implement a δ -function in the probability distribution of the CB mass ratio. Such an equal mass ratio population is also evident in our colour–magnitude diagram (dashed grey lines in Fig. 2), with our analysis suggesting that the equal mass likelihood $P_{\text{eqm}} = 0.04$ (see Appendix A). The origin of such a twin population is unclear, but it could be related to gas accretion onto a forming binary star tending to equalize the masses (Tokovinin 2017; Tokovinin & Moe 2020; Tokovinin 2023). To include this,

Table 1. The r_{sky} and \tilde{v} range used for the WBT and the binning scheme, which we fixed in advance to limit the possibility of bias (Banik, Pittordis & Sutherland 2021).

r_{sky} (kAU)	\tilde{v}
2–5 (6 bins)	0–1.6 (20 bins)
5–10 (5 bins)	1.6–2.6 (5 bins)
10–16 (3 bins)	2.6–5 (5 bins)
16–22 (2 bins)	
22–30 (2 bins)	

we combine the CB perturbations calculated assuming a smooth distribution of the CB mass ratio with similar calculations assuming an exactly equal mass ratio for the CB, weighting the former at $(1 - P_{\text{eqm}})$ and the latter at P_{eqm} (further details are provided in Banik, Pittordis & Sutherland 2021). We demonstrate later that changing the assumed CB mass ratio distribution has very little impact on our results (Section 5.1.2).

The overall impact of CBs on the WB population is modulated by the likelihood f_{CB} that a star in our WB sample has an undetected CB companion. We assume that CB contamination of the two stars in each WB occurs independently, so the likelihood of both stars being contaminated is f_{CB}^2 , which is not negligible. We found that one of the most computationally expensive parts of the WBT is calculating the CB perturbations to the WB parameters in this double contaminated case. This is due to the need to consider a range of CB semi-major axes, velocity perturbations, and mass contributions from both CBs, which almost doubles the already large number of nested loops.

Combining the CB perturbations with the WB library is non-trivial because CBs cannot be considered as adding an extra population to the statistics – they physically affect the WB. As summarized in equation (25), we need to scale up the WB \tilde{v} to account for the fractional mass increase $\Delta\tilde{M}$ induced by the CB(s) and then add the velocity perturbations from each CB in the directions parallel and orthogonal to \tilde{v} of the WB, which is a 2D vector in the sky plane. We then use the Pythagoras Rule to calculate the resultant \tilde{v} , or more precisely the appropriate \tilde{v} pixel (Table 1). Since there is a reasonable chance that the resultant $\tilde{v} > 5$, CBs can reduce the overall normalization of the $(r_{\text{sky}}, \tilde{v})$ distribution in our WB library. This may well have happened in reality, but since our *Gaia* DR3 sample only contains WBs with $\tilde{v} < 5$, we renormalize the $(r_{\text{sky}}, \tilde{v})$ distribution that we obtain after the CB-WB convolution step. The computational cost of convolving the CB and WB libraries is kept low using the techniques detailed in Banik, Pittordis & Sutherland (2021).

3.3 Chance alignments

So far we have only considered bound systems. It is inevitable that some of the WBs in our sample will be unbound. Observations are unlikely to catch a WB in the process of disruption due to a passing third star or a molecular cloud (see section 8.1 of BZ18). Even so, chance alignments of stars can still arise, especially when we get to large r_{sky} and \tilde{v} (see fig. 7 of PS19). The likelihood is enhanced due to the possibility of the stars being born in the same star cluster which later dissolved [Oh et al. 2017; Dinnbier, Kroupa & Anderson 2022], giving them a much lower relative velocity than the Galactic velocity dispersion. This raises the chance of observing the stars close to each other on the sky, which could cause them to be misidentified as a WB given the *Gaia* parallax uncertainty. LOS contamination should be

reduced somewhat by excluding regions of the sky with an enhanced density of stars, e.g. known star clusters (table 1 of PS23).

To limit the complexity of our analysis, we set up a very simple model for LOS contamination under the assumption that the separation and relative velocity of stars born in the same cluster can greatly exceed that of the WBs in our sample. Since \mathbf{r}_{sky} and \mathbf{v}_{sky} are both 2D vectors within the sky plane, we assume that the number density of chance-aligned WBs scales as

$$\frac{dN_{\text{LOS}}}{dr_{\text{sky}} dv_{\text{sky}}} \propto r_{\text{sky}} v_{\text{sky}}. \quad (26)$$

Bearing in mind that $v_{\text{sky}} \propto \tilde{v}/\sqrt{r_{\text{sky}}}$ [equation (3)], the population distribution of LOS contamination is

$$\frac{dN_{\text{LOS}}}{dr_{\text{sky}} d\tilde{v}} \propto \tilde{v}. \quad (27)$$

The dependence on r_{sky} cancels out because raising r_{sky} reduces the range of \mathbf{v}_{sky} for the same range in \tilde{v} .

We integrate this distribution across each of the r_{sky} and \tilde{v} pixels shown in Table 1 to obtain the relative number of chance-aligned systems that we expect to find in each pixel. We then normalize the results to a sum of f_{LOS} and combine with the result of our CB-WB convolution (Section 3.2.4) weighted by $(1 - f_{\text{LOS}})$, where f_{LOS} is the LOS contamination fraction. We leave f_{LOS} as a free parameter in our analysis. Since there are only a few WBs in our catalogue at high r_{sky} and \tilde{v} where LOS contamination would be most significant (Fig. 6), we expect f_{LOS} to be at most a few percent.

3.4 The simulated distribution of WBs

To simulate the distribution of WBs, we tie together a model for the WB population (Section 3.1) with a model for an undetected population of CB companions (Section 3.2) and then allow for chance alignments (Section 3.3). We use Fig. 10 to illustrate the separate stages for the particular example of our best-fitting MOND model (Section 4.1). The simulated \tilde{v} distribution is shown in four different r_{sky} ranges considering only WBs (blue curves), including CBs (red curves), and finally including also the LOS contamination (black curves). Since only a small proportion of the WB sample is expected to lie in any single r_{sky} interval used here, we normalize the \tilde{v} distributions using the final simulated distribution, so only the black curve in each panel is guaranteed to have an integral of 1. WBs alone almost never yield $\tilde{v} \gtrsim 1.5$, even in MOND – the theoretical limit of 1.7 only refers to systems on highly eccentric orbits seen close to pericentre from a direction nearly orthogonal to the relative velocity (this is also apparent in fig. 3 of BZ18). CB companions create an extended declining tail that is easily able to reach much higher \tilde{v} . We capture much of this extended tail by going up to $\tilde{v} = 5$, a choice we discussed in Section 2.1 and fixed in advance of the WBT to mitigate possible biases (Banik, Pittordis & Sutherland 2021). This extra leverage allows our analysis to constrain the CB population, which dominates the behaviour once $\tilde{v} \gtrsim 2$. A flat log-space distribution of a_{int} (Öpik 1924) yields a \tilde{v} distribution that is also flat in log-space in the CB-dominated region, helping to explain why the distribution here is $\sim 1/\tilde{v}$. We expect this to flatten out and ultimately start rising at very high \tilde{v} due to LOS contamination. This is most evident in the highest r_{sky} interval, where there is less contribution from only the gravitationally bound systems because WBs have a declining distribution of r_{sky} [equation (16)]. Our WB sample reaches high enough \tilde{v} for its distribution to become flat, but a clearly rising trend is not apparent because we have limited the considered \tilde{v} range to avoid being swamped by chance alignments.

Other studies which consider a wider parameter range do show a rising trend (e.g. see fig. 12 of El-Badry, Rix & Heintz 2021).

3.5 Comparison with observations

We compare the observed number of WBs in each $(r_{\text{sky}}, \tilde{v})$ pixel with the expectation of our model for that pixel. We use binomial statistics because we are dealing with integer statistics and the total sample size N is finite. Thus, the binomial likelihood of observing k systems in a pixel is

$$P_{\text{pixel}} = \frac{N!}{(N-k)!k!} p^k (1-p)^{N-k}, \quad (28)$$

where p is the fraction of the total number of WBs in the comparison region ($r_{\text{sky}} = 2\text{--}30$ kAU, $\tilde{v} < 5$) which should be located within the pixel under consideration. For computational reasons, we use the logarithmic version of this equation. The log-likelihood $\ln P$ of the model is then given by considering all 540 pixels summarized in Table 1.

$$\ln P = \sum_{\text{Pixels}} \ln P_{\text{pixel}}. \quad (29)$$

This is used to infer the optimal values and confidence intervals for our model parameters associated with the CBs (f_{CB} , k_{CB} , and γ), LOS contamination (f_{LOS}), and the WBs (a_{break} , β , and α_{grav}). Note that γ affects both the CB and WB populations because they are assumed to follow the same eccentricity distribution for simplicity. The CBs are assumed to follow an Öpik law distribution of semi-major axes (Öpik 1924) over a 1.5 dex range (Section 3.2.3), while the WBs follow a broken power law that reduces to $a^{-1.6}$ at small separations [equation (16)].

Notice that regardless of how many WBs a pixel contains, $\ln P_{\text{pixel}} < 0$ in every pixel as its predicted p is always non-zero once we include CB and LOS contamination. Since we have quite a large number of WBs spread across a large number of pixels, we expect $\ln P \ll 0$ (a similar situation is evident in table 2 of Asencio et al. 2022). As a crude estimate, we can suppose that our model predicts $8611/540 = 16$ WBs in each of our 540 pixels, so the actual number of WBs in any pixel is expected to nearly follow a Poisson distribution with a standard deviation of $\sqrt{16} = 4$. If the model works well, then the actual number of WBs would be approximately uniformly distributed over the range 12–20, with all of these outcomes having a likelihood of $P_{\text{pixel}} = 1/9$. The combined likelihood would then be $\ln P = -540 \ln 9 = -1200$. We therefore expect to get a combined log-likelihood of about this much, but this is of course only a very crude estimate. In reality, not all pixels would be expected to have the same number of WBs, and even if they were, outcomes in the tail of the Poisson distribution are less likely than the inverse of the Poisson noise. Moreover, it is inevitable that our model does not capture all the subtleties of the actual WB population because of the simplifying assumptions made to keep the complexity and computational cost manageable. None the less, we will see later that the above estimate is fairly accurate.

4 RESULTS

Before presenting the results of our detailed statistical analysis discussed in the previous section, we first look for trends in the median \tilde{v} of our WB population with r_{sky}/r_M , which is a proxy for the internal acceleration of each WB. For this purpose, we sort our WBs in order of r_{sky}/r_M and use this to create ten equally sized subsamples with no overlap. The thin magenta line in Fig. 11 shows

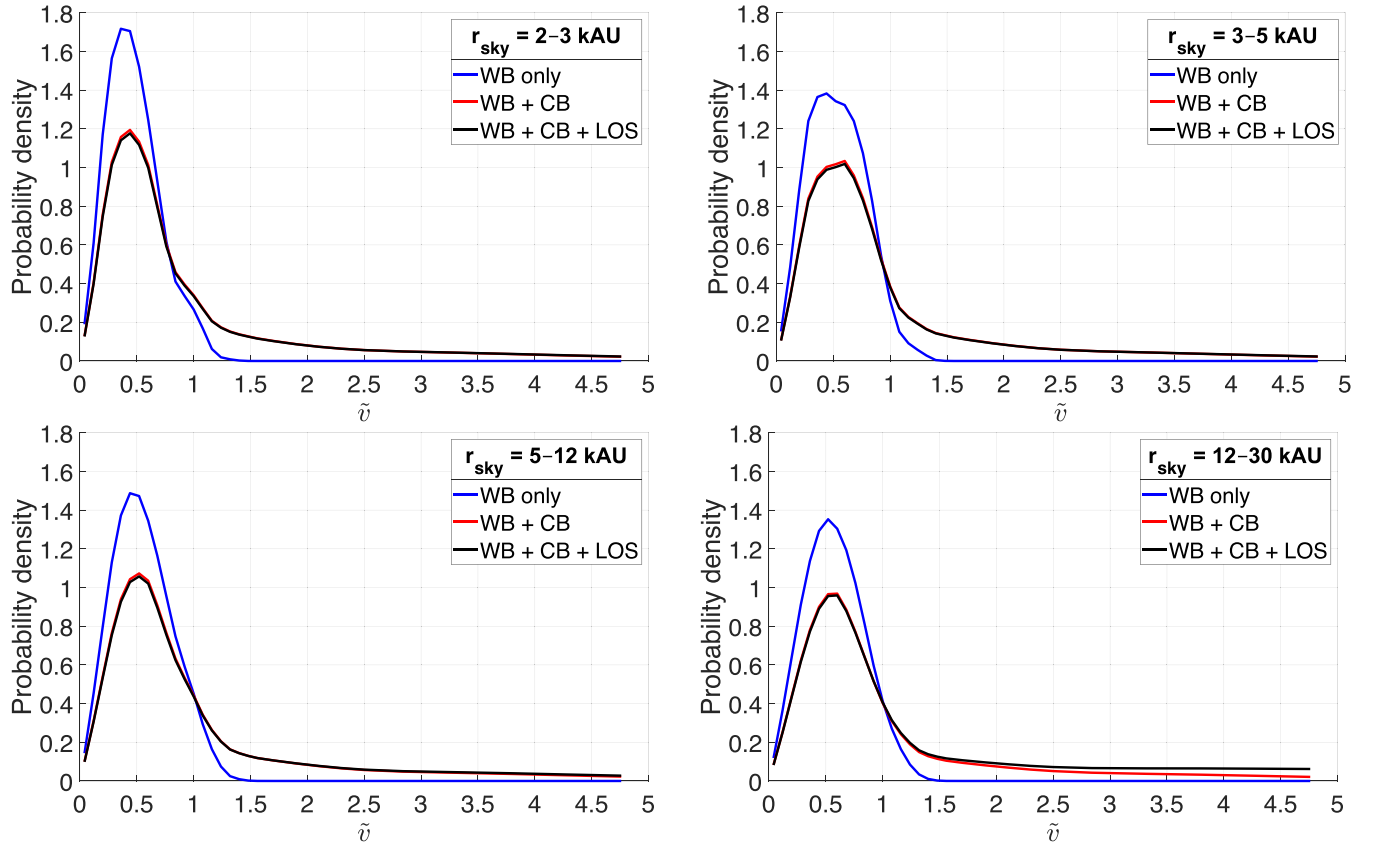


Figure 10. How we build up the simulated \tilde{v} distribution in our best MOND model (Section 4.1), with results shown in four different r_{sky} ranges (different panels). In each case, the distribution due to WBs alone (blue) is significantly altered by including CBs (red). Further adding LOS contamination yields our final simulated distribution (black), which is used to normalize the results shown here. This helps to highlight that including LOS contamination slightly raises the predicted number of WBs at high r_{sky} .

the median \tilde{v} as a function of the median r_{sky}/r_M of each subsample. A clear rising trend is evident, with the increase being close to the 20 per cent expected in MOND. However, LOS contamination would be more important at high r_{sky} and high \tilde{v} , potentially inflating the median \tilde{v} in a manner which appears acceleration-dependent. To mitigate against this possibility, we note that MOND would cause an acceleration-dependent broadening to the main peak of the \tilde{v} distribution at $\tilde{v} \lesssim 2$ (Fig. 6; see also fig. 7 of PS19). In MOND, we expect almost no WBs with $\tilde{v} > 1.5$ in the absence of contaminating effects (see fig. 3 of BZ18). We therefore consider the median \tilde{v} of only those WBs in each r_{sky}/r_M bin with $\tilde{v} < 1.5$, 2, or 2.5, which we show in Fig. 11 using the red, black, and blue line, respectively. In all three cases, the rising trend disappears: the median \tilde{v} becomes flat with respect to our proxy for the internal acceleration. There is of course some scatter, but we would expect a random scatter of about $0.6/\sqrt{861} = 0.02$ due to the finite number of WBs, which is roughly in line with the actual scatter.

Since the upper limit to \tilde{v} should substantially alleviate the impact of CB and LOS contamination without much affecting a genuine MOND signal, it is clear that the apparent MOND signal discussed above is almost certainly a consequence of contaminating effects, highlighting how easily one can reach erroneous conclusions from the WBT using simple population statistics without a good understanding of the astrophysical systematics (Hernandez et al. 2019; McCulloch & Lucio 2019; Hernandez, Cookson & Cortés 2022). Focusing on WBs with $\tilde{v} \lesssim 2$, the very weak dependence of

the median \tilde{v} on r_{sky}/r_M despite the significant range covered by our sample strongly suggests that the WBT will return a Newtonian result. We highlight this using the dashed grey lines in Fig. 11, which approximately show the MOND prediction that the median \tilde{v} should rise by a factor of $\sqrt{\eta} = 1.2$ over the range $r_{\text{sky}}/r_M = 0.2 - 1$. This is motivated by Fig. 7, which shows that the enhancement factor to the radial Newtonian gravity rises from 1 up to its asymptotic value roughly linearly over this range of r/r_M and is almost flat afterwards. While this is a rather crude way of considering the situation because it neglects projection and orbital phase effects, it should approximately capture the expected broadening of the \tilde{v} distribution in the main peak region as one considers WB subsamples with a lower internal acceleration. The expected Newtonian behaviour is not shown as there is already a gridline at 0.57. This provides a much better fit to the median \tilde{v} for the $\tilde{v} < 2$ subsample than the approximate MOND expectation mentioned above.

This strong hint that the WBT will disfavour MOND is in line with the detailed analysis of PS23, who prepared a model for the WB population similar to ours and found a strong preference for Newtonian gravity over MOND. Their Newtonian model fits the observations surprisingly well given their limited exploration of the parameter space (see their fig. 12). Moreover, the right-hand panel of their fig. 17 shows that WBs have a median $v_{\text{sky}} \propto 1/\sqrt{r_{\text{sky}}}$, especially when restricting to only those WBs with $\tilde{v} \lesssim 2$. This is the expected behaviour if Kepler's Third Law works even in the low-acceleration regime, as illustrated by the dashed lines on this figure.

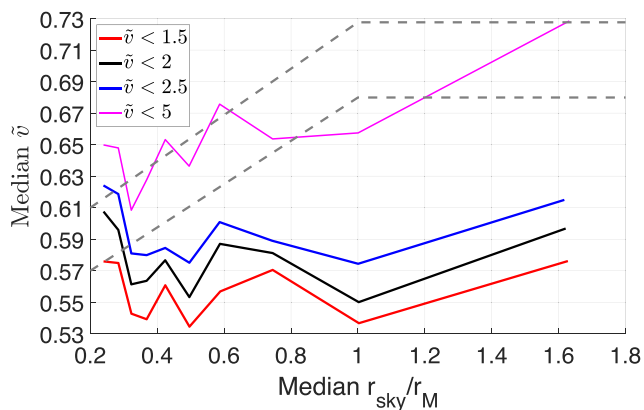


Figure 11. The median \tilde{v} of our WB sample as a function of the median r_{sky}/r_M , shown for ten bins in the latter which each contain 861 WBs except the last bin which has 862. The solid lines show results with the four different upper limits to \tilde{v} of 1.5 (red), 2 (black), 2.5 (blue), and 5 (thin magenta). Qualitatively, the MOND expectation is that the median \tilde{v} rises for larger r_{sky}/r_M and then becomes flat once the Galactic gravity dominates. We indicate this by two dashed grey lines with different normalizations. Their shape is based on Fig. 7, which implies that the main peak in the \tilde{v} distribution for $\tilde{v} \lesssim 2$ should broaden at low accelerations. Neglecting projection effects and assuming the current separation is representative of the whole orbit, this would lead to a rise in the median \tilde{v} by a factor of $\sqrt{\eta} = 1.2$ that occurs roughly linearly over the range $r_{\text{sky}}/r_M = 0.2 - 1$ before rapidly flattening out. The Newtonian prediction is a flat relation, which is consistent with the data up to $\tilde{v} < 2.5$, within which lies 89.8 per cent of the sample. The full sample of WBs (with $\tilde{v} < 5$) does show a rising median that looks qualitatively similar to the MOND expectation, but since this rise is not seen for the subsamples restricted to the main peak region of the \tilde{v} distribution where \tilde{v} is low enough to be physically plausible for an uncontaminated system, we attribute this behaviour to LOS contamination becoming more important at larger separations (Section 3.3).

The lack of a trend in the median \tilde{v} is unlikely to be caused by observational uncertainties, which would generally lead to a larger \tilde{v} uncertainty at high r_{sky} due to the lower Newtonian v_c [equation (3)]. This would if anything create the appearance of a MOND-like signal in actually Newtonian data rather than precisely cancel out a genuine MOND signal (cf. Section 5.2). Moreover, the \tilde{v} uncertainties in our WB sample are very small (Fig. 4).

Although our results favour the Newtonian model, the distribution of \tilde{v} contains much more information than just its median value. Before drawing strong conclusions about the gravity law, we try to exploit this information using a forward model that considers the most relevant factors (Section 3). We now turn to the results of a comparison between this model and the WB data set.

4.1 Gradient ascent with a fixed gravity law

To check if Newtonian gravity indeed provides a better fit to the data than MOND, we fix the gravity law to Newtonian ($\alpha_{\text{grav}} = 0$) and run a gradient ascent in the remaining model parameters to try and maximise $\ln P$ (Fletcher & Powell 1963). The gradient ascent usually takes about 100 iterations to converge. To further improve the fit, we supplement this by running a ‘line ascent’ where $\ln P$ is maximized with respect to each parameter while holding all the other parameters fixed. We then repeat this extended version of gradient ascent for MOND ($\alpha_{\text{grav}} = 1$). The parameters of the best-fitting models are shown in Table 2, whose last line shows the corresponding $\ln P$. This is much lower in the MOND case, with $\Delta \ln P = 175$. This can be

Table 2. The best-fitting model parameters with a fixed gravity law, found using gradient ascent. Notice the very substantial difference in $\ln P$ (Section 3.5), indicating that MOND provides a much poorer fit to the data.

Parameter	Gravity law	
	Newton	MOND
a_{break} (kAU)	5.10	5.30
β	-2.66	-2.70
γ	1.86	1.96
f_{CB} (%)	69.9	65.7
k_{CB} (%)	1.07	2.10
f_{LOS} (%)	1.45	1.49
α_{grav} (fixed)	0	1
$\ln P$	-1457.4	-1632.6

thought of as equivalent to a statistical significance of $\sqrt{2\Delta \ln P}$ standard deviations for a single Gaussian variable, implying that the WBT prefers Newtonian gravity over MOND at 19σ confidence. The high significance is in line with earlier forecasts that a few thousand systems should be more than sufficient for the WBT (see the blue curves in fig. 5 of BZ18).

Our result broadly agrees with that of PS23 if we consider their Newtonian and MOND models with a thermal eccentricity distribution ($\gamma = 1$), which is the closest distribution they consider to the slightly superthermal distribution preferred by our fits. Their analysis gives a difference in χ^2 between the best Newtonian and MOND models of 525, implying that Newtonian gravity is preferred over MOND at a significance of $\sqrt{525}\sigma$ or about 23σ . We expect that the actual significance would be somewhat lower because further exploration of the parameter space should improve the fit, with MOND benefitting far more due to its $2.6\times$ higher χ^2 . Even so, it is clear that the WBT favours Newtonian gravity over MOND at $\gg 5\sigma$ confidence. We will see later that this conclusion holds up in our more thorough exploration of the parameter space. Other attempts at the WBT are discussed in Section 5.2.

To better understand why the Newtonian fit to the WB data set is much better than the MOND fit, we use Fig. 12 to plot the \tilde{v} distribution in four different r_{sky} ranges, normalized using the observed number of systems in each r_{sky} range. The error bars show binomial uncertainties in the model predictions. The top left panel has little direct sensitivity to the gravity law because of the low r_{sky} , but there is an indirect dependence because the best model parameters differ depending on the gravity law. The lower f_{CB} in the best MOND fit (solid blue line) causes the \tilde{v} distribution to peak earlier at a much higher amplitude and then drop off more rapidly after the peak compared to the best Newtonian fit (solid black line). This causes MOND to severely disagree with the data, which may seem surprising as MOND is not directly expected to be very relevant at such low r_{sky} . The disagreement must arise because of a tradeoff when simultaneously fitting the data over a wide r_{sky} range, so we turn to the situation at wider separations. As the gravity law gradually becomes more important, the predicted mode to the Milgromian \tilde{v} distribution eventually ‘overtakes’ the Newtonian prediction.¹¹ The Newtonian and MOND distributions become almost identical at $r_{\text{sky}} = 3-5$ kAU (top right panel). The Newtonian model performs slightly better around the peak of the distribution because it predicts an earlier

¹¹Fig. 3 of BZ18 shows how the eccentricity distribution and the gravity law affect the \tilde{v} distribution, but here we focus on optimized fits for each gravity theory to enable a fairer comparison.

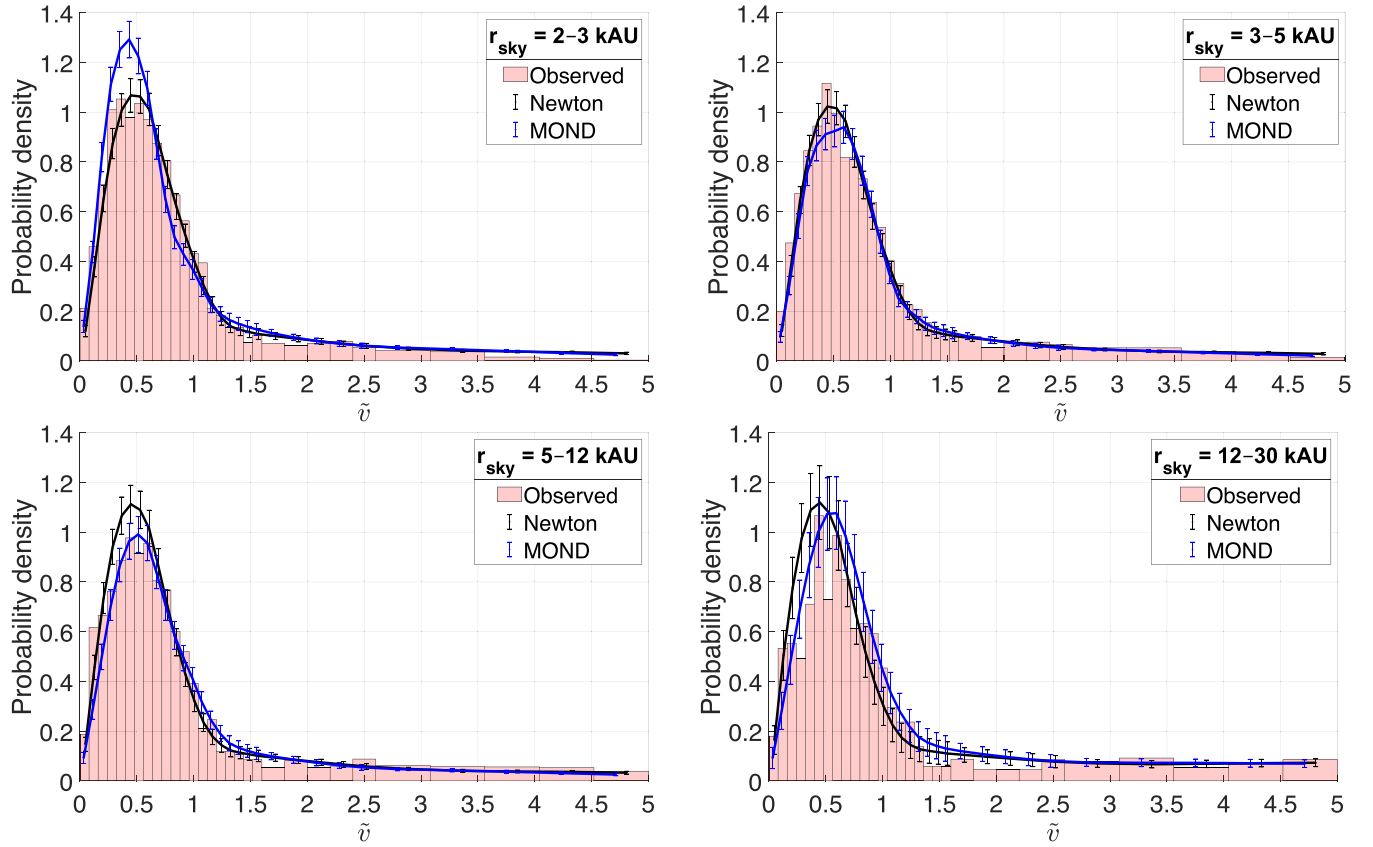


Figure 12. Comparison between the observed \tilde{v} distribution in four different r_{sky} ranges (different panels) with the prediction of our best Newtonian and Milgromian model (black and blue lines with error bars, respectively). Results with Newtonian (Milgromian) gravity are offset very slightly to the right (left) from the centre of each \tilde{v} bin for clarity in case the predictions are similar. Uncertainties are calculated using binomial statistics. The results shown here are normalized based on the observed WB distribution, so the simulated curves in each panel need not have an integral of 1.

and higher peak despite the higher f_{CB} , highlighting that the gravity law is now one of the dominant factors. This remains true at larger separations, where MOND always predicts a later peak to the \tilde{v} distribution. For $r_{\text{sky}} = 5 - 12$ kAU (bottom left panel), MOND performs somewhat better in a handful of pixels around the peak region, though given the uncertainties, the Newtonian model is not that far off. The rapid decline after the peak and the very low number counts at $\tilde{v} = 1 - 1.5$ work better in the Newtonian model. The situation is similar for $r_{\text{sky}} = 12 - 30$ kAU (bottom right panel), where MOND again fits slightly better before the peak in the \tilde{v} distribution. The uncertainties are now discernibly larger due to the smaller proportion of the total WB sample expected (and observed) to be in this range compared to the other r_{sky} ranges mentioned above.¹² Even so, the peak position is better reproduced in Newtonian gravity, which also provides a somewhat better match to the sharp observed decline afterwards over the range $0.5 - 1.5$ and the consequent rather low number counts over the range $1 - 1.5$. This range was expected to be critical for the WBT (section 5 of BZ18) and is also a region where the model predictions depend almost entirely on the gravity law, which is obviously not the case near the peak of the distribution where other factors are also relevant (see their fig. 3). Both models provide a very good fit to the almost flat \tilde{v} distribution over the range $2.5 - 5$, nicely demonstrating a balance between a declining trend

from the extended CB tail and a rising trend from LOS contamination [equation (27)]. The rapid observed decline just beyond the peak of the \tilde{v} distribution when $r_{\text{sky}} \gtrsim 3$ kAU presumably forces the MOND fit to use a lower f_{CB} , but this also causes the MOND model to provide a worse fit to precisely this range of \tilde{v} at very low r_{sky} , as shown in the top left panel. Indeed, the combination of a too-rapid decline after the main peak at very low r_{sky} with a too-gradual decline at high r_{sky} is just what we would expect of an optimized Milgromian fit if the predicted change to the shape of the \tilde{v} distribution does not occur in the real data, as strongly suggested by Fig. 11.

It is of course inevitable that Newtonian gravity would fit better in some pixels and MOND in others, so it is important to consider the overall goodness of fit [equation (29)]. To better pinpoint which regions of parameter space contribute to the much lower likelihood in MOND, we find the difference in $\ln P$ between the best-fitting Newtonian and MOND models for each pixel and sum the results across one of the dimensions. Fig. 13 shows the sum of the so-obtained $\Delta \ln P_{\text{pixel}}$ values for all \tilde{v} pixels at fixed r_{sky} (left-hand panel) and vice versa (right-hand panel). The left-hand panel shows that all r_{sky} pixels below 12 kAU show a preference for Newtonian gravity, which is sometimes very strongly preferred. In particular, the range $r_{\text{sky}} = 5 - 12$ kAU provides the bulk of the evidence in favour of Newtonian gravity, even though we saw earlier that MOND works slightly better around the peak of the \tilde{v} distribution at these separations. Beyond 12 kAU, 4/6 of the r_{sky} pixels prefer MOND, but the preference is very weak in all cases. The right-hand panel shows that nearly all \tilde{v} pixels prefer Newtonian gravity. The handful of \tilde{v}

¹²Binomial uncertainties in the model prediction for each pixel do not depend on how many WBs are observed in it.

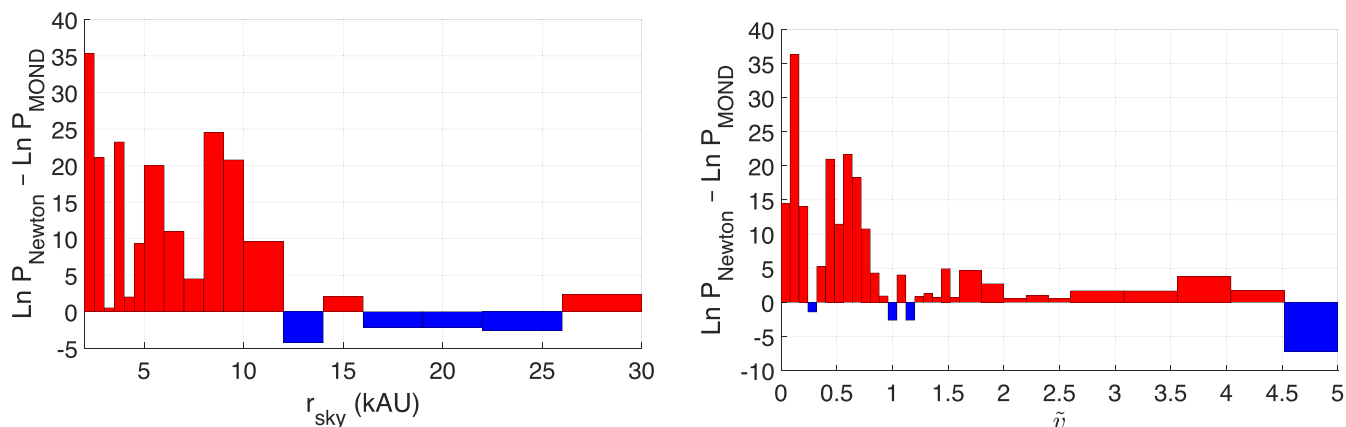


Figure 13. The difference in log-likelihood between our best Newtonian and Milgromian models, shown here after summing across all pixels in \tilde{v} at each r_{sky} (left-hand panel) and vice versa (right-hand panel). Blue bars indicate that MOND works better. Most bars are red, indicating a better fit in Newtonian gravity. Overall, $\ln P$ is higher in the Newtonian model by $\Delta \ln P = 175.2$, which suggests a preference for Newtonian gravity over MOND by $\sqrt{2\Delta \ln P}$ standard deviations, i.e. at 18.7σ confidence. We use Fig. C1 to show a pixel-by-pixel comparison between the log-likelihoods returned by each model.

pixels which prefer MOND appear to be randomly distributed. The mild preference for MOND at very high \tilde{v} could be driven by a deficit of WBs here at very low r_{sky} due to our sample selection imposing that $v_{\text{sky}} < 3 \text{ km s}^{-1}$ (Section 2.1). At our lowest considered r_{sky} of 2 kAU and assuming $M = M_{\odot}$, this limit corresponds to $\tilde{v} < 4.5$, which would cause only a very mild edge effect considering the very low number of WBs that we might reasonably expect in the affected pixels. This issue is further mitigated in a revised analysis where we use a narrower range of M (Section 5.1.1). We use Fig. C1 to present the 2D version of the results shown here, with a colour scheme used to show all the $\Delta \ln P_{\text{pixel}}$ values and small white circles to highlight pixels which work better in MOND, as sometimes occurs to a very limited extent.

4.2 Markov Chain Monte Carlo (MCMC) analysis

Our main results are based on running an MCMC analysis starting from the optimal parameters identified by gradient ascent, which we start at $\alpha_{\text{grav}} = 1/2$ to avoid biasing the results in favour of either gravity theory we are testing.¹³ MCMC is a standard statistical method that generates a sequence of parameter values whose frequency distribution matches the posterior inference on the model parameters. Starting from some initial guess for the parameters with likelihood P , the protocol calls for generating a proposal by adding Gaussian random perturbations to the parameters. The revised parameters lead to a model with likelihood P_{next} . We follow the Metropolis-Hastings approach to MCMC in which the proposal is accepted if $P_{\text{next}} > P$, but if not, then it is accepted with a likelihood of only P_{next}/P by using a random number generator. Every time the proposal is accepted, the parameter perturbations are applied and P is updated. When the proposal is rejected, the previous parameters must be recorded again. For best results, the acceptance fraction should be close to 23.4 per cent (Gelman, Gilks & Roberts 1997). We ensure it is always in the range 21 per cent – 26 per cent, which should ensure a well-mixed chain. This is achieved by setting the

¹³Using gradient ascent to initialize an MCMC analysis worked well in Asencio et al. (2022), though here we supplement the gradient ascent with a line ascent stage where only one parameter is varied at a time, with the gravity law being the last parameter to be optimized prior to the MCMC due to its importance.

perturbation on each parameter to have a Gaussian dispersion similar to the estimated uncertainty, which we guess using our gradient ascent algorithm based on the curvature of $\ln P$ with respect to that parameter in the supplementary stage when all the other parameters are fixed (Section 4.1). We then ran a reduced length MCMC chain to estimate the parameter uncertainties more precisely and set the Gaussian dispersion of each parameter to its estimated uncertainty scaled by a common factor close to 2/3. This factor cannot be varied inside an MCMC chain, so we ran a few reduced length chains to find out that this is the appropriate factor to use in our nominal analysis (slightly different factors were needed in the variations discussed in Section 5.1).¹⁴ All the MCMC analyses presented in this contribution use 10^5 trials.

Fig. 14 shows the result of our MCMC analysis as a ‘triangle plot’ (Bocquet & Carter 2016), with the result for our nominal assumptions shown in red in all such plots. This assumes a uniform prior on the eccentricity index γ [equation (17)] over the range 0 – 4. The blue lines show the results assuming instead a Gaussian prior of $\gamma = 1.32 \pm 0.09$ (Hwang, Ting & Zakamska 2022). The top panel in each column shows the inference on just one model parameter marginalized over all others. The other panels show the inference on every pair of parameters, which clarifies whether there are any correlations in the uncertainties. There are only two noticeable cases of correlated errors, both of which are anti-correlations. The first of these is between the parameters governing the semi-major axis distribution of the WBs [equation (16)]. This is similar to the correlation that can arise between the slope and intercept of a linear regression if the data are clustered about a point offset from the origin (a similar correlation is evident between the parameters r_{core} and Slope_r in fig. 10 of Asencio et al. 2022). The other correlation evident from Fig. 14 is between f_{CB} and k_{CB} , which are the only two parameters that relate solely to the CB population. Reducing k_{CB} causes the CBs to orbit faster and have a larger impact on \tilde{v} , spreading out the CB tail over a larger range in \tilde{v} and thus reducing its amplitude. To match the observed amplitude of the extended tail to the \tilde{v} distribution in the regime where neither genuine WB orbital motion nor LOS contamination should be significant, the algorithm is forced to increase the prevalence of CB companions.

¹⁴An alternative approach is to let the factor vary dynamically and then use the optimal value in a full length chain.

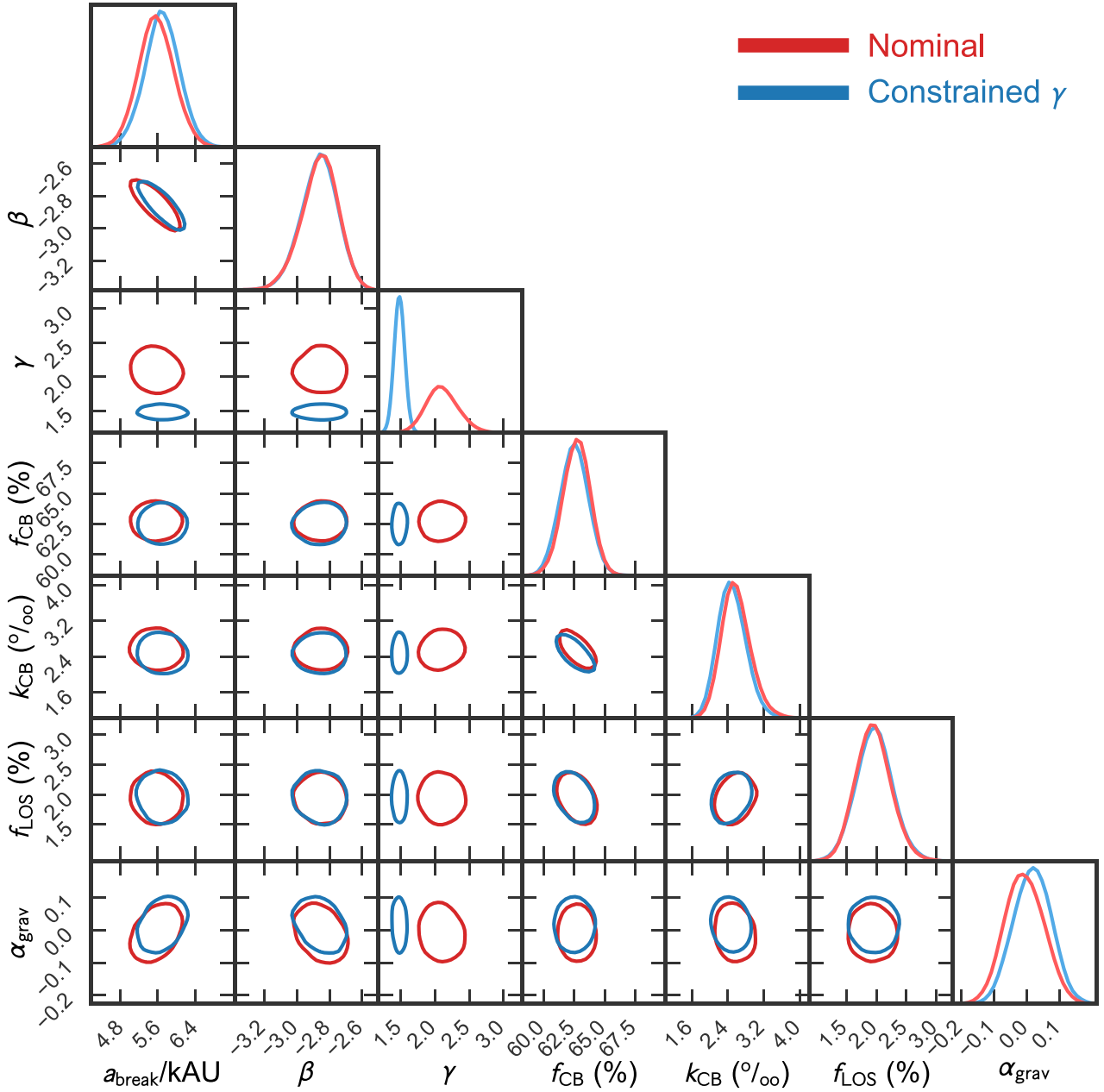


Figure 14. Triangle plot showing the posterior inference on each model parameter (top panels in each column) and the 68 per cent confidence region for every parameter pair (other panels). The red lines show our nominal analysis with a wide uninformative prior on γ [equation (17)], while the blue lines impose a prior of $\gamma = 1.32 \pm 0.09$ based on the angle between the sky-projected separation and relative velocity of each WB (Hwang, Ting & Zakamska 2022). Notice how this constraint has little effect on the other model parameters, including especially the gravity law, which is clearly Newtonian (bottom right panel). All the triangle plots in this contribution were prepared using PYGTC (Bocquet & Carter 2016) and show the nominal analysis with solid red lines.

Applying a restrictive prior on γ has barely any effect on the other parameter inferences, though of course the inference on γ is considerably tighter in this case. We see that in our nominal analysis where γ is allowed to vary freely, it prefers a somewhat higher value than 1.32. This could indicate that the WB orbital eccentricity distribution is more super-thermal than reported by Hwang, Ting & Zakamska (2022). We argued previously that this is quite possible given that CB and LOS contamination were not considered in their analysis but would tend to drive the distribution of the angle ψ towards uniform, which is the expected angular distribution if $\gamma = 1$.

The preferred values of the parameters related to the WB semi-major axis distribution (a_{break} and β) are roughly in line with the results of Andrews, Chanamé & Agüeros (2017), with the steeper decline at high a likely due to our quality cuts selecting against large separation WBs where it can be harder to get tight constraints on \tilde{v} . The occurrence rate of CBs (related to the parameter f_{CB}) is slightly higher than the 50 per cent estimated by Clarke (2020). However, stars do appear to have binary companions quite frequently, with observations indicating a fraction exceeding 40 per cent (Hartman, Lépine & Medan 2022) and usually only placing a lower limit because the observations are only sensitive to some range of orbital

Table 3. The posterior inference on each of our model parameters, showing its mode and 1σ confidence interval. We provide a reference to the equation or section where the model parameter is defined. The first row with values shows the uniform prior on each parameter for our nominal analysis. Each subsequent row shows the result of an MCMC analysis, beginning with the nominal assumptions. Revised analyses are discussed further in Section 5.1. The last column shows the model with the highest likelihood for each set of assumptions, considering both the MCMC chain and the final result of the gradient ascent used to initialize it.

Altered assumption	Model parameter							Best $\ln P$
	a_{break} (kAU)	β	γ	f_{CB} (%)	k_{CB} (‰)	f_{LOS} (%)	α_{grav}	
Definition	equation (16)	equation (16)	equation (17)	Section 3.2.4	Section 3.2.3	Section 3.3	Section 3.1.3	Section 3.5
Prior	(1, 15)	(−15, −1)	(0, 4)	(0, 99)	(1, 750)	(0, 60)	(−2, 3.6)	–
Nominal	$5.61^{+0.33}_{-0.42}$	$-2.84^{+0.09}_{-0.12}$	$2.04^{+0.27}_{-0.19}$	$62.81^{+1.05}_{-1.20}$	$2.52^{+0.32}_{-0.27}$	$1.94^{+0.29}_{-0.27}$	$-0.021^{+0.065}_{-0.045}$	−1438.98
$\gamma = 1.32 \pm 0.09$	$5.64^{+0.42}_{-0.32}$	$-2.84^{+0.09}_{-0.13}$	$1.47^{+0.08}_{-0.07}$	$62.54^{+1.06}_{-1.12}$	$2.37^{+0.37}_{-0.20}$	$1.99^{+0.26}_{-0.31}$	$0.024^{+0.042}_{-0.054}$	−1445.58
$M = (1 - 2)M_{\odot}$	$5.78^{+0.47}_{-0.32}$	$-3.07^{+0.13}_{-0.14}$	$1.86^{+0.30}_{-0.18}$	$63.52^{+1.01}_{-1.45}$	$2.43^{+0.26}_{-0.32}$	$1.75^{+0.35}_{-0.31}$	$-0.038^{+0.076}_{-0.052}$	−1329.82
$M_{\star} > 0.5 M_{\odot}$	$7.13^{+0.92}_{-0.79}$	$-2.63^{+0.14}_{-0.24}$	$1.49^{+0.36}_{-0.09}$	$61.82^{+1.25}_{-1.27}$	$3.93^{+0.62}_{-0.55}$	$1.82^{+0.49}_{-0.33}$	$-0.177^{+0.064}_{-0.081}$	−1263.68
Flat q for CBs	$5.53^{+0.37}_{-0.39}$	$-2.84^{+0.11}_{-0.12}$	$2.15^{+0.25}_{-0.25}$	$62.94^{+0.98}_{-0.99}$	$2.43^{+0.28}_{-0.30}$	$2.00^{+0.29}_{-0.30}$	$-0.033^{+0.074}_{-0.056}$	−1441.28
Linear q for CBs	$5.66^{+0.28}_{-0.47}$	$-2.88^{+0.13}_{-0.07}$	$2.06^{+0.24}_{-0.21}$	$62.78^{+1.07}_{-1.10}$	$2.47^{+0.27}_{-0.32}$	$1.99^{+0.22}_{-0.37}$	$-0.006^{+0.057}_{-0.056}$	−1441.18
$k_{\text{CB}} = 0.2$	$3.84^{+0.22}_{-0.30}$	$-2.73^{+0.07}_{-0.09}$	$4.00^{+0.00}_{-0.14}$	$67.08^{+1.52}_{-1.62}$	200	$9.38^{+0.48}_{-0.36}$	$-0.271^{+0.071}_{-0.090}$	−1994.27
Tri 1.5 dex	$5.51^{+0.37}_{-0.37}$	$-2.83^{+0.07}_{-0.12}$	$2.12^{+0.24}_{-0.20}$	$59.51^{+1.34}_{-1.21}$	$3.40^{+0.45}_{-0.52}$	$1.94^{+0.32}_{-0.28}$	$-0.016^{+0.049}_{-0.061}$	−1434.16
Flat 1 dex	$5.60^{+0.37}_{-0.35}$	$-2.87^{+0.11}_{-0.10}$	$1.99^{+0.19}_{-0.24}$	$56.51^{+1.17}_{-1.01}$	$2.05^{+0.26}_{-0.18}$	$1.95^{+0.30}_{-0.28}$	$0.017^{+0.041}_{-0.066}$	−1443.14
$f_{\text{CB}} = 0.3$	$5.65^{+0.27}_{-0.38}$	$-2.97^{+0.09}_{-0.13}$	$1.05^{+0.09}_{-0.14}$	30	$18.7^{+1.8}_{-0.9}$	$7.33^{+0.47}_{-0.31}$	$0.149^{+0.035}_{-0.048}$	−1852.05

parameters (Riddle et al. 2015). After correcting for incompleteness, fig. 1 of Offner et al. (2023) suggests that a reasonable f_{CB} for our sample might be 50 per cent if most CB companions are not identified in our faint star catalogue (Section 2.1) or through our cut on the parameter $ipd_frac_multi_peak$ (Section 2.4.3). We will see later that we can reduce our inferred f_{CB} from our nominal 63 ± 1 per cent to about this much with some slight changes to the CB model that have little impact on the other parameters (Section 5.1.3). The inferred k_{CB} of $2.5 \pm 0.3\%$ is very much in line with our estimate based on the various selection effects at play (Section 3.2.3), even though our prior merely imposed that $k_{\text{CB}} = 1 - 750\%$. Only a few percent of the WBs in our sample appear to be LOS contamination, which was to be expected from the WB distribution shown in Fig. 6 – though it is also clear that we cannot adequately fit the data without this ingredient.

Since the non-gravitational parameters seem broadly in line with what one might expect, we now turn to the inferred gravity law. This is very consistent with Newtonian ($\alpha_{\text{grav}} = 0$) and completely rules out MOND ($\alpha_{\text{grav}} = 1$). A clear Newtonian result is not surprising in light of the model-independent Fig. 11, our discussion in Section 4.1, and the results of PS23 using a somewhat different data set and model for the WB population (comparisons with their study and other attempts at the WBT are discussed in Section 5.2).

Our prior for each model parameter is shown in Table 3 along with the mode and 1σ confidence region of its posterior inference in all of our MCMC analyses (different rows). The alterations to the modelling assumptions are discussed in Section 5.1. The last column shows the log-likelihood of the model where this is maximal, considering both the final result of the gradient ascent and the whole MCMC chain. Only the analysis with f_{CB} fixed at a rather low value prefers $\alpha_{\text{grav}} > 0$ at any reasonable significance. However, this model has an extremely low log-likelihood compared to our nominal assumptions, suggesting a 29σ preference for the latter (the very severe problems with this model are discussed further in Appendix D). Thus, it is not realistically possible for the WBT to give a Milgromian result. We explain later why constraints from galaxy rotation curves preclude altering the MOND interpolating function so as to evade our constraint from the WBT (Section 5.3).

5 DISCUSSION

Since the main goal of this contribution is to distinguish Newtonian gravity from MOND, our *a priori* expectation was that α_{grav} should be 0 or 1. Our results strongly prefer 0. If local WBs are actually Milgromian, there would have to be a confluence of observational errors and modelling deficiencies that shift α_{grav} from a true value of ≈ 1 down to almost exactly 0. Measurement errors on \tilde{v} are very small in our WB sample (Fig. 4) compared to the intrinsic dispersion of ≈ 0.5 (Fig. 12). Moreover, uncertainties would tend to broaden the distribution, with the broadening occurring preferentially at high r_{sky} because the same velocity uncertainty in m s^{-1} translates into a larger uncertainty in \tilde{v} [equation (3)]. Thus, measurement errors would if anything cause a Newtonian WB population to look somewhat Milgromian. This is the opposite of what it would take to reconcile our results with MOND.

The impact of modelling deficiencies is harder to assess, but PS23 implemented the WBT with a somewhat different sample selection and model for the WB population. Those authors considered WBs with $r_{\text{sky}} = 5\text{--}20$ kAU rather than our adopted range of $2\text{--}30$ kAU, leading to a smaller proportion of systems deep into the Newtonian regime. To get a better handle on the CB properties, they considered systems with $\tilde{v} < 7$ rather than $\tilde{v} < 5$. In addition, they used a linear mass-luminosity relation rather than our cubic fit over the range most relevant to the WBT (Fig. 1). Despite these and other differences, they also found a very strong preference for Newtonian gravity over MOND, with the Newtonian model fitting the observed number counts rather well (fig. 12 of PS23). Their analysis with a thermal eccentricity distribution ($\gamma = 1$) provides the best fit in both gravity theories and is also closest to the somewhat superthermal distribution preferred by our fits. It also better matches the distribution of the angle between the sky-projected separation and relative velocity of each WB (Hwang, Ting & Zakamska 2022). For the case $\gamma = 1$, the analysis of PS23 prefers Newtonian gravity over MOND at about 23σ confidence. This is similar to our result in Section 4.1 that the best Newtonian model is preferred over the best MOND model at 19σ confidence.

An important reason for the similar results is that the gravity law is mostly sensitive to the main peak region of the \tilde{v} distribution, so

it is not necessary to have a very accurate model for the extended tail at $\tilde{v} \gtrsim 2$. This does not of course mean that we can completely neglect CB and LOS contamination, but it does mean that the manner in which these effects are included should have little effect on the inferred gravity law. Even differences in the eccentricity distribution have a less significant effect on the overall \tilde{v} distribution than whether gravity is Newtonian or Milgromian (see fig. 3 of BZ18). This is also evident in Fig. 14, which shows that forcing γ to a lower value than preferred in our nominal analysis only marginally affects the inferred α_{grav} and its uncertainty. There is almost no correlation between these parameters in their joint inference.

We have assumed that γ is independent of other parameters like the WB semi-major axis. A correlation here can masquerade as a correlation with the WB orbital acceleration because g_N depends more sensitively on the separation than it does on M and the dynamic range in r_{sky} is larger than in the mass. It is possible that WBs with smaller separations mostly formed together, while more widely separated WBs formed via capture of field stars. This could lead to WBs with larger a being on typically more eccentric orbits [higher γ in equation (17)], which would reduce the median \tilde{v} because the WB would spend more time near apocentre (see fig. 3 of BZ18). In principle, this could counteract the MOND prediction that WBs in the asymptotic regime have enhanced \tilde{v} due to the predicted enhancement to the gravity binding the WB. However, it seems unlikely that the cancellation would be sufficiently precise to yield the observed very flat trend of the median \tilde{v} with respect to r_{sky}/r_M (Fig. 11).

To further explore this quite contrived possibility, we compute the \tilde{v} distribution for a grid of Keplerian orbits with different eccentricities. For each orbit, we consider all possible orbital phases and viewing angles.¹⁵ We then assume some eccentricity distribution governed by the parameter γ [equation (17)]. The median of the resulting \tilde{v} distribution depends on the assumed γ , as shown by the red line on Fig. 15. We add a horizontal solid blue line showing the median \tilde{v} of 0.64 when $\gamma = 1$. Bearing in mind that γ affects the distribution of the angle ψ between \mathbf{r}_{rel} and \mathbf{v}_{rel} , this is the lowest plausible γ for Newtonian WBs with separations of only a few kAU (see fig. 7 of Hwang, Ting & Zakamska 2022). Since lower γ corresponds to a higher median \tilde{v} , 0.64 is a conservative upper limit to the median \tilde{v} of our Newtonian WBs. This is also evident from the horizontal grey band showing the median \tilde{v} of the WB population, which is not that high (see also Fig. 11). To mask a MOND signal, WBs in the asymptotic regime would need to have a median \tilde{v} which is smaller by a factor of $\sqrt{\eta}$ due solely to changes in the eccentricity distribution. We illustrate this with a dashed blue line at $0.64/\sqrt{\eta} = 0.53$. It is not possible to reach such a low median \tilde{v} even if $\gamma = 4$, which would significantly affect the observed distribution of ψ (see fig. 2 of Hwang, Ting & Zakamska 2022). Using this constraint, their fig. 7 shows that there is no tendency for WBs at larger separations to prefer systematically higher γ (the last three bins are most relevant for the WBT). Their fig. 10 also shows that for a slightly superthermal eccentricity distribution (γ slightly above 1), their method of determining γ would preferentially overestimate it at large separations, making it even less likely that $\gamma \gg 1$ in this regime. Moreover, the actual distribution of ψ in their fig. 6 appears fairly symmetric with respect to $\psi \rightarrow \pi - \psi$ for WBs with $r_{\text{sky}} > 1$ kAU, providing an important sanity check. The observed approximate symmetry between systems heading towards and away from each other makes it unlikely that our results are contaminated

¹⁵The computational costs are reduced by only considering the period between pericentre and apocentre.

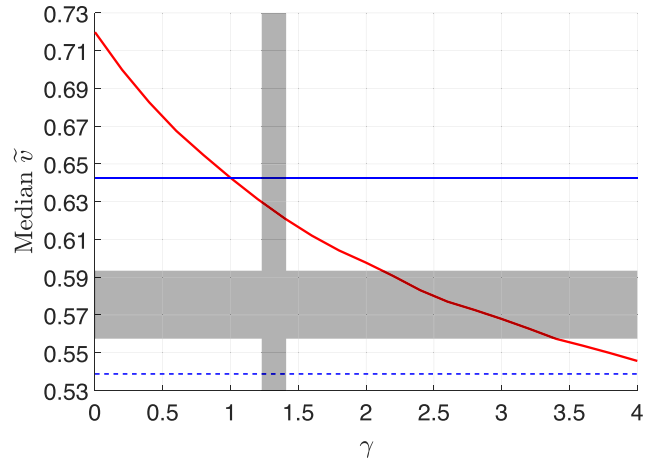


Figure 15. The median \tilde{v} for an ensemble of Keplerian orbits with a range of eccentricities viewed from a random viewing angle at a random orbital phase (red line). The eccentricity distribution is governed by γ [equation (17)]. The horizontal solid blue line shows the result when $\gamma = 1$. To mask a MOND signal, the median \tilde{v} would need to be reduced by a factor of $\sqrt{\eta}$ to the level shown by the horizontal dashed blue line. The vertical grey shaded band shows $\gamma = 1.32 \pm 0.09$ based on observations of WBs with separations similar to those we use for the WBT (Section 3.1.2; see also Hwang, Ting & Zakamska 2022). The horizontal grey shaded band shows the median \tilde{v} of the WB population, with the range giving the uncertainty arising from whether we truncate the \tilde{v} distribution at 1.5, 2, or 2.5 when finding the median. The WBs are not further binned in r_{sky}/r_M as this has little impact on the median \tilde{v} (Fig. 11).

much by recently ionized WBs that are currently dispersing. This is not too surprising given that unbound systems are expected to disperse quickly and that disruptive encounters with perturbers like passing stars would be rather infrequent (see section 8.1 of BZ18). It is therefore extremely unlikely that a rapidly rising trend in γ with a can explain the very clear Newtonian result obtained by our analysis and that of PS23.

Another possible systematic effect is that WBs with larger separations could preferentially be younger because they are more fragile to disruption, which could systematically affect the mass-luminosity relation. However, it has been estimated that the half-life $t_{1/2}$ of a WB in the Solar neighbourhood is comparable to the 10 Gyr age of the Galactic disc (Knox, Hawkins & Hambly 1999) only when $a = 31$ kAU (equation 1 of Jiang & Tremaine 2010). Survival rates of WBs would be reduced modestly in MOND because although their binding energy is enhanced by the gravity boost factor η , the impulse from a passing star is also enhanced by the same factor, so the energy gained per interaction would be enhanced by a factor of η^2 if we treat impulses from different sources as contributing to a diffusive random walk (heating rates from impulsive encounters in MOND are discussed extensively in section 3.3.2 of Asencio et al. 2022). This would reduce $t_{1/2}$ by a factor of η , but since $t_{1/2} \propto a^{-2}$ (equation 4 of Bahcall, Hut & Tremaine 1985), we can maintain the value of $t_{1/2}$ by reducing a . This would entail reducing the above estimate of 31 kAU by a factor of $\sqrt{\eta}$, which would reduce it to 26 kAU. Even taking this into account, there should only be a mild tendency for WBs to preferentially be younger at larger separations when $r_{\text{sky}} < 10$ kAU, which is the range that mostly contributes to the WBT preferring Newtonian gravity over MOND (left-hand panel of Fig. 13). WBs in the Solar neighbourhood must be fairly robust given the large number identified within 250 pc (PS23) and the larger number within 1 kpc (El-Badry, Rix & Heintz 2021). In

fact, the nearest star to the Sun is in a WB (Kervella, Thévenin & Lovis 2017), with Monte Carlo simulations of this system indicating a 74 per cent survival probability over 5 Gyr (Feng & Jones 2018).

Even if there is some correlation between the separation of a WB and the ages of its stars, it is unclear how that would influence the estimated mass. Main sequence stars increase their brightness over time as their cores become denser and hotter due to the fusion of hydrogen into helium, so younger stars would tend to be less luminous (Hejlesen 1980). This would cause their mass to be underestimated and their \tilde{v} to be overestimated [equation (3)]. If this occurs more commonly at large separations, it would enhance rather than hide a MOND signal in the data. However, stars that are not yet on the main sequence become fainter as they settle onto it, so younger stars would be brighter at this evolutionary stage (Hayashi 1961; Hayashi & Hoshi 1961). Depending on the time required to settle onto the main sequence, it could be that the effect is important at the low mass end, though we note that most of the stars in our WB sample have $M_* > 0.3 M_\odot$ (blue bars in Fig. 5). We further mitigate this issue by explicitly considering the impact of removing all WBs where either star has $M_* < 0.5 M_\odot$ (Section 5.1.1).

We estimate the mass of each star from its apparent G -band magnitude and trigonometric parallax. While these should both be reliable, we also need to assume a relation between the absolute G -band magnitude and mass. For this, we fit a polynomial which is easily inverted, thereby smoothing over small-scale features (Section 2.2). One such feature is the Jao gap (Jao et al. 2018), which arises because stars with $M_* \lesssim 0.35 M_\odot$ have fully convective cores (Chabrier & Baraffe 2000). This leads to an unstable range of luminosities marked by the convective kissing instability (Mansfield & Kroupa 2021). However, the gap has a width of only 0.05 magnitudes (section 2 of Jao et al. 2018). This has a discernible impact on the luminosity function of stars (fig. 17 of Gaia Collaboration 2021a), but the impact on the mass–luminosity relation is very small because 0.05 mag corresponds to only a 1 per cent difference in mass [equation (7)] and thus a 0.5 per cent difference in \tilde{v} [equation (3)]. A more significant issue might be the broad hump evident in the mass–luminosity relation at $M_* \approx 0.5 M_\odot$ (Kroupa, Tout & Gilmore 1993; Reid, Gizis & Hawley 2002), which is probably the same feature evident at $M_G \approx 10$ in fig. 16 of Gaia Collaboration (2021a). While the mass–luminosity relation here is important to our analysis (blue bars in Fig. 5), a smooth polynomial fit to the V -band mass–luminosity relation only deviates from the data points by a few tenths of a magnitude (fig. 11 of Reid, Gizis & Hawley 2002). With accurate data, gaps in the mass–luminosity relation or sharp changes in its gradient can lead to sharp features in the distribution of absolute magnitudes (as with the Jao gap), but the steep mass–luminosity relation mitigates the impact on the estimated M_* . Moreover, a polynomial fit to the mass–luminosity relation should not deviate systematically from the actual relation: if the fit overestimates the mass at some M_G , then it should underestimate the mass at a slightly different M_G . In any case, such modest deviations from our multipart polynomial fit should be much less substantial than the difference between the cubic fit we use over most of the relevant range and the linear fit adopted by PS23 in their equation 3. Since there is a very strong preference for Newtonian gravity over MOND in both studies and these even give a similar statistical significance (Section 4.1), it is unlikely that this preference is caused by issues with the assumed mass–luminosity relation, which moreover is calibrated with spectroscopic mass estimates from FLAME (Section 2.3.3).

Another important aspect of the WBT is the presence of undetected CB companions, which broaden the tail of the \tilde{v} distribution (Fig. 10). We have assumed that every star in our WB sample is equally likely

to have such a companion. However, it is possible that f_{CB} depends on properties of the star and even its WB companion. For this to affect the WBT, f_{CB} would need to depend on the WB acceleration, which mostly depends on its separation. Counteracting the effect of MOND would require more widely separated WBs to have a smaller f_{CB} , which could hold down the median \tilde{v} in the main peak region and hide the MOND signal. This seems unlikely because the CBs relevant to our analysis have a much smaller separation than the WB (Section 3.2.3), something that is also clear from the fact that our inferred $k_{CB} \ll 1$. We would therefore expect f_{CB} to not depend very much on the separation of the WB (as also suggested by spectroscopic searches for binaries; Tokovinin & Smekhov 2002; Tokovinin, Hartung & Hayward 2010).

To test this expectation, we need a proxy for f_{CB} . We develop one based on the fact that CBs affect the tail of the \tilde{v} distribution. Even in MOND gravity, we expect WBs to very rarely have $\tilde{v} > 1.5$ in the absence of contaminating effects (Fig. 10). This motivates us to use the proportion of WBs with higher \tilde{v} as a measure of contamination. While this arises from both CBs and chance alignments, Fig. 6 shows that the latter is not very important if we restrict to $\tilde{v} \lesssim 2.5$. This is because LOS contamination becomes more important at high \tilde{v} [equation (27)], opposite the situation for CB contamination (PS19). We therefore argue that the incidence of CBs can be gauged by finding the likelihood P_{tail} that a WB with $\tilde{v} < 2.5$ also has $\tilde{v} > 1.5$.

$$P_{\text{tail}} \equiv P(\tilde{v} > 1.5 | \tilde{v} < 2.5). \quad (30)$$

While P_{tail} cannot be equated with f_{CB} , a higher incidence of CBs would increase P_{tail} , making it a useful proxy for the prevalence of CB contamination. We therefore quantify P_{tail} and its uncertainty using binomial statistics. Assuming a uniform prior on P_{tail} , its posterior distribution is characterized by (equation 32 of Asencio et al. 2022):

$$\begin{aligned} \text{mean} &= \frac{N_{>1.5} + 1}{N_{<2.5} + 2}, \\ \text{standard deviation} &= \frac{1}{N_{<2.5} + 2} \sqrt{\frac{(N_{>1.5} + 1)(N_{<2.5} - N_{>1.5} + 1)}{N_{<2.5} + 3}}, \end{aligned} \quad (31)$$

where $N_{<2.5}$ is the number of WBs with $\tilde{v} < 2.5$, of which $N_{>1.5}$ WBs also have $\tilde{v} > 1.5$. If $N_{>1.5} = N_{<2.5} = 0$, the posterior distribution of P_{tail} is uniform over the range 0 – 1 and we obtain the expected result for its mean and standard deviation.

To check if P_{tail} depends on the WB separation, we determine P_{tail} for the WBs in four r_{sky} intervals (the same as those used in Fig. 12). The result is shown in Fig. 16. There is no discernible trend in P_{tail} : a value of 8 per cent can adequately fit all four data points despite the rather small binomial uncertainties. This makes it very unlikely that our analysis mistakenly prefers Newtonian gravity due to trends in the likelihood of a star having a CB companion.

We test the robustness of our falsification of MOND by varying some of our modelling assumptions (Section 5.1) and considering WBT results from other authors (Section 5.2). We then consider if uncertainties in the MOND interpolating function can plausibly allow WBs in the asymptotic regime to have orbital velocities that deviate only a few percent from the Newtonian prediction, as indicated by the WBT (Section 5.3). We will see that it is not possible to reconcile MOND with the WBT through plausible variations to the modelling assumptions or the interpolating function. We go on to discuss the broader implications of this result, focusing on whether the WB data and the tight observed RAR (Li et al. 2018; Desmond 2023) can be explained by further adjustments to MOND or by a completely new theory, bearing in mind other astrophysical constraints like Solar System ephemerides and the RAR in galaxy clusters (Section 5.4).

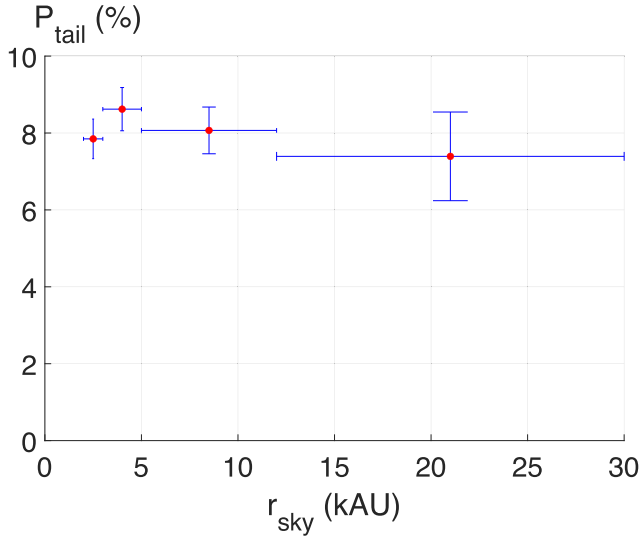


Figure 16. The impact of r_{sky} on P_{tail} [equation (30)], which we argue in the text should correlate with the CB fraction and thus serves as a useful proxy for the latter. P_{tail} is shown for WBs in the same four r_{sky} bins as those used in Fig. 12. Error bars in P_{tail} show binomial uncertainties [equation (31)], while error bars in r_{sky} show the width of each bin.

A Newtonian result from the WBT causes severe difficulties for any theory which approximately reduces to the MOND field equation on kAU scales in the Solar neighbourhood, forcing us to consider if some other approach might be necessary.

5.1 Variations to the nominal analysis

In this section, we vary some of the modelling assumptions in our nominal analysis. The posterior inferences on the model parameters are summarized in Table 3.

5.1.1 Restricted mass range for the WB stars

Our model assumes that several quantities like the CB contamination fraction f_{CB} are independent of the WB mass M . It could be that some of these quantities do in fact depend on M . Moreover, higher mass stars are more luminous, allowing us to detect them out to a greater distance given the requirement that the apparent magnitude $m_G < 17$ (Section 2.1). Since the total mass of a WB affects its MOND radius and thus the predicted enhancement to its Newtonian gravity, it is possible that our predicted \tilde{v} distribution in different r_{sky} bins is somewhat inaccurate due to using the overall distribution of M as a prior in all r_{sky} bins.

While it is not feasible to use different M distributions in different r_{sky} bins due to the increased complexity and computational cost, we can use a narrower mass range to mitigate modelling deficiencies of this kind. Since this reduces the sample size, we roughly centre our restricted mass range on the mode of our WB sample’s mass distribution (red bars in Fig. 5). This leads us to consider only WBs with a mass of $1.0 - 2.0 M_{\odot}$. Since the impact of CBs depends on the distribution of M_{\star} for the stars in our sample, we need to recalculate the CB velocity perturbations with revised $\Delta\tilde{M}$ and f_{pb} (equations 18 and 20, respectively). We also adjust the distribution of M used to marginalise our WB orbit library and rerun our analysis in Appendix A, which now gives $P_{\text{eqm}} = 0.03$. The results are shown in brown in the top left panel of Fig. 17. The preferred gravity law

is now even further from MOND, though it remains very consistent with the Newtonian prediction and differs only modestly from our nominal analysis. Shifts to the other parameter inferences are also small. Since the restricted mass range substantially reduces the scope for changes to the mass distribution across the parameter range used for the WBT, our results appear to be robust in this respect.

The WBT relies on having an accurate estimate of M_{\star} for every star in our WB sample. Our M_{\star} estimates are based on the mass-luminosity relation from Pecaut & Mamajek (2013), but their accuracy has been improved slightly by calibrating against spectroscopic estimates from the *Gaia* FLAME package (Section 2.3.3). This calibration is only possible when $M_{\star} > 0.5 M_{\odot}$ as there is essentially no data below this. We therefore consider restricting our WB sample to only those systems where both stars have $M_{\star} > 0.5 M_{\odot}$, which reduces P_{eqm} to 0.01. The results are shown in blue in the top left panel of Fig. 17. Apart from a $1.6\times$ increased k_{CB} which we discuss below, the most noteworthy change is a reduction in α_{grav} to slightly negative values, causing almost 3σ tension with the Newtonian prediction ($\alpha_{\text{grav}} = 0$). This shift in the inferred gravity law could indicate that the higher typical M_{\star} of the sample increases the lowest a_{int} that we need to consider for the CBs. Our estimate for this in Appendix B assumes that the M_{\star} distribution easily extends down to $0.4 M_{\odot}$, which is correct for our nominal sample (blue bars in Fig. 5) but is no longer true if we require $M_{\star} > 0.5 M_{\odot}$. A higher mass would increase the separation corresponding to the same orbital period. Since this does not affect our argument in Section 3.2.3 about the upper limit to the relevant range of a_{int} , the dynamic range of a_{int} might be less than our nominal assumption of 1.5 dex. Keeping it fixed at 1.5 dex to minimize the changes could then force the analysis to raise the upper limit to a_{int} by increasing k_{CB} , which does indeed happen (see Table 3). We note that in a different extension analysis, reducing the allowed range of a_{int} to 1 dex increases the inferred α_{grav} by 0.04 (Section 5.1.3). A similar change to the result of our $M_{\star} > 0.5 M_{\odot}$ analysis would reduce the discrepancy with Newtonian gravity to about 2σ .

5.1.2 The CB mass ratio distribution

We assume that the CB mass ratio distribution $\propto q^{0.4}$ ($q \leq 1$) based on Korntreff, Kaczmarek & Pfalzner (2012). We showed in Fig. 9 that this provides a good match to our WB sample, especially when bearing in mind that selection effects can make very low q systems hard to detect. Since the CBs are undetected, it would also be reasonable to consider a flat prior on q or a linear prior, both of which seem marginally plausible. Considering these revised priors on q should thus bracket the uncertainty in its distribution. In particular, a flat distribution roughly approximates the three-part power law recommended in fig. 7 of El-Badry et al. (2019).

The results of these two revised analyses are shown in the top right panel of Fig. 17, with the linear prior shown in blue and the flat prior shown in brown. While the actual distribution of q is inevitably more complicated than the power-law forms considered in this study, the posterior inferences of all seven model parameters and the overall goodness of fit barely change in both analysis variants (Table 3). Our results thus seem robust to what exactly we assume here.

5.1.3 The CB semi-major axis distribution

The impact of CBs on the WB \tilde{v} depends on the distribution of a_{int} , so we consider variations to our assumptions regarding it. We would generally expect any CB to become unstable if $k_{\text{CB}} \gtrsim 0.3$ (PS23), but our analysis prefers a value of only 2.5%. We argued

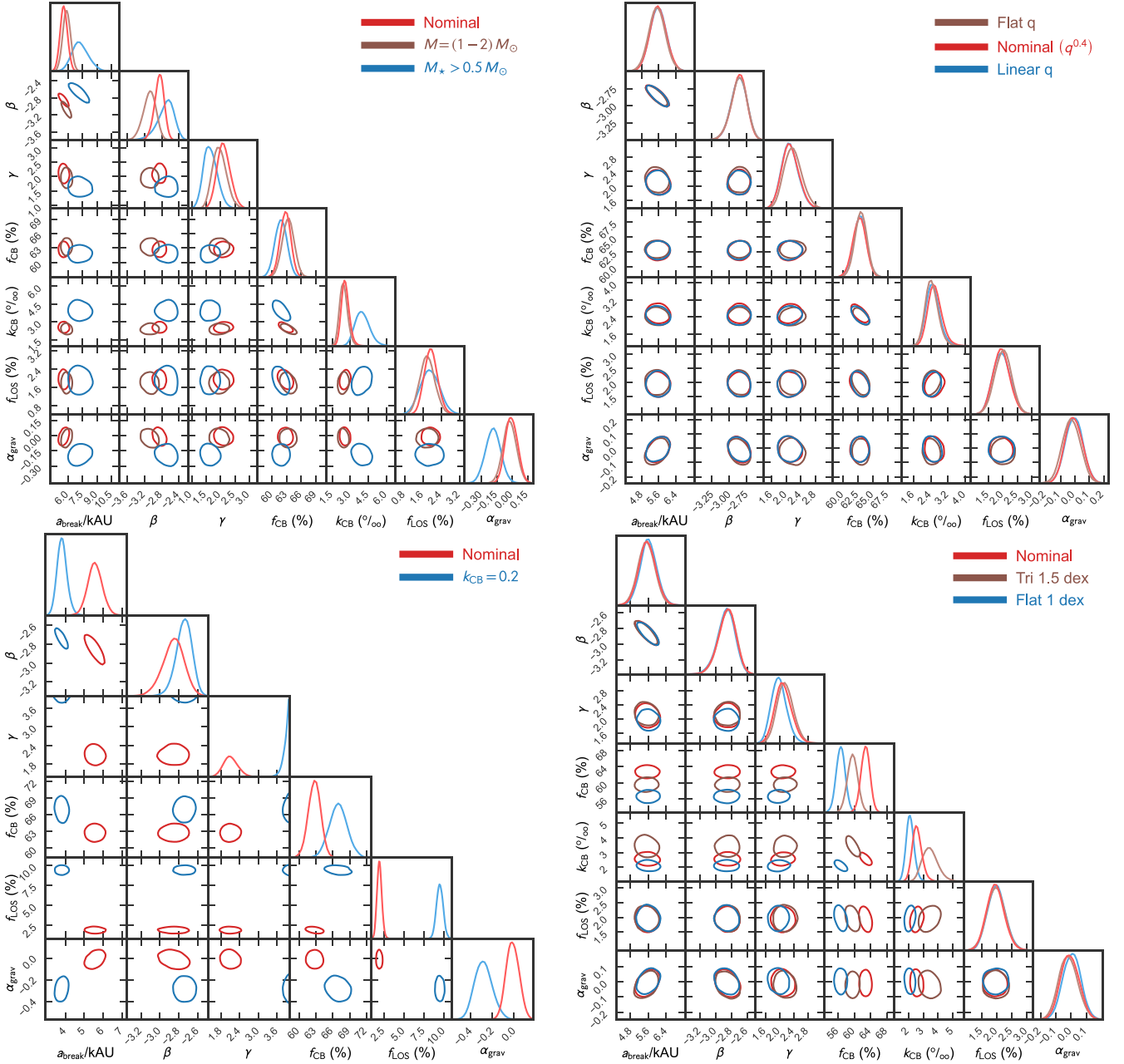


Figure 17. Triangle plot similar to Fig. 14, but now comparing the nominal analysis (solid red lines in all panels) to revised analyses where the total mass of each WB is restricted to the range $1.0 - 2.0 M_{\odot}$ or the individual stars in each WB are required to have $M_{\star} > 0.5 M_{\odot}$ (top left; Section 5.1.1), the CB mass ratio distribution is altered to bracket the uncertainties as illustrated in Fig. 9 (top right; Section 5.1.2), the maximum semi-major axis of the CBs is fixed at 20 per cent of the WB semi-major axis (bottom left; Section 5.1.3), and the logarithm of the CB semi-major axis is assumed to have a triangular distribution of width 1.5 dex or a flat distribution of width 1 dex (bottom right; Section 5.1.3). The posterior inferences on the model parameters in these revised analyses are compared with the nominal analysis in Table 3.

that this is to be expected given the various selection effects at play as part of our strategy to minimize CB contamination and model the remaining CBs (Section 3.2.3). In particular, the faint star catalogue (Section 2.1) and the use of the *Gaia* catalogue parameter *ipd_frac_multi_peak* (Section 2.4.3) should remove CBs based on their outright or marginal detection, respectively.

Even so, it is worth considering what happens if we fix $k_{\text{CB}} = 0.2$ in equation (22). The results are shown in the bottom left panel of Fig. 17, but with the rows and columns for k_{CB} omitted as this is now fixed. Compared to our nominal analysis where k_{CB} is a free

parameter, the fits are pushed towards unrealistically high values of γ , contradicting the results of Hwang, Ting & Zakamska (2022). This could be due to the much higher k_{CB} reducing the velocity perturbations from CBs and making it harder to fit the extended tail to the \tilde{v} distribution. This problem can be slightly alleviated by postulating higher orbital eccentricities, a higher fraction of CBs, and a higher fraction of chance-aligned WBs – all of which are evident from the posteriors. The inferred f_{LOS} of 9.4 per cent is particularly high compared to the 1.9 per cent in our nominal analysis. It is hard to imagine that the small number of WBs towards the upper right

of Fig. 6 is consistent with 9 per cent of the full sample of WBs being chance alignments, which would be less common at lower \tilde{v} [equation (27)]. These issues are reflected in the much poorer overall fit by a factor of $\exp(555)$; last column in Table 3). Even if we leave aside these very serious issues, the inferred gravity law is further from Milgromian than in our nominal analysis to such an extent that it is now barely consistent with Newtonian, indicating that a higher k_{CB} cannot reconcile MOND with the WBT.

We also expect our results to depend somewhat on the width of the a_{int} distribution [equation (22)]. We have assumed a flat logarithmic distribution for simplicity. We might expect the distribution to decline to zero more gradually at the edges. We therefore also consider a triangular logarithmic distribution in which the mode occurs at the centre of the considered range, i.e. 0.75 dex from either end. This extension analysis is labelled ‘Tri 1.5 dex’ in Table 3, where we see that the fit improves slightly compared to our nominal analysis. The only noticeable shift to the parameter inferences is a slight decrease to f_{CB} and corresponding increase to k_{CB} , roughly parallel to the direction along which their errors are anti-correlated (bottom right panel of Fig. 17). Importantly, the change to the inferred gravity law is about $10\times$ smaller than its uncertainty.

Our nominal analysis assumes a 1.5 dex width in a_{int} for reasons discussed in Section 3.2.3. To check how much this assumption affects our results, we now reduce this to 1 dex while continuing to assume a flat logarithmic distribution. This extension analysis is labelled ‘Flat 1 dex’ in Table 3, where we see that the fit becomes slightly worse than in our nominal analysis. The only parameter shift that is much larger than the random uncertainty is a significant reduction in f_{CB} , which is reduced from 63 per cent to 57 per cent. Despite significantly narrowing the allowed range of a_{int} , there is only a modest reduction to the inferred k_{CB} , bolstering our argument in Section 3.2.3 that it should be about 2.5%. The inferred gravity law also barely changes. On the other hand, the reduced incidence of CBs suggests that it is difficult for our model to reliably estimate just how common they actually are.

This difficulty probably involves CBs with low separations that cause large \tilde{v} perturbations, leading to a much broader low amplitude tail to the \tilde{v} distribution. Since we only consider WBs with $\tilde{v} < 5$, it is possible that such tight CBs actually matter less for our analysis. In that scenario, there would be a degeneracy between the frequency of these very tight CBs and the overall CB fraction. Roughly speaking, suppose that we have a CB model where 50 per cent of the CBs create such large velocity perturbations that any WB containing such a subsystem would very likely have $\tilde{v} > 5$ and thus be rejected from our sample, while the remaining CBs create more palatable velocity perturbations. If our analysis then needs 30 per cent of the WBs in our sample to contain a CB in order to adequately fit the extended tail at $\tilde{v} \gtrsim 1.5$, then we would have to assume that the likelihood of a WB containing a CB is about 46 per cent. This is because half of the contaminated systems would exit our sample altogether rather than contribute to the extended \tilde{v} tail, so the proportion of the remaining WBs which contain a CB would be about $23/(100 - 23) = 0.3$. However, we can envisage a different model in which none of the CBs are very tight, so any instance of CB contamination is very unlikely to cause the WB to have $\tilde{v} > 5$. In this case, it would be possible to make do with a lower overall incidence of CBs while still adequately explaining the extended \tilde{v} tail. While this would inevitably affect the detailed shape of the distribution, this could be hard to tell given that our model has other degrees of freedom related to the WB population and the extent of LOS contamination. Moreover, our motivation for truncating the a_{int} distribution at the low end is related to the short orbital period (Appendix B) rather than

the velocity perturbations becoming so large that the contaminated WB is likely to exit our sample, an effect that is already included in our analysis. It is beyond the scope of our model to directly consider the astrometric oscillations that must arise with very tight CBs and how they would impact \tilde{v} . These issues might be mitigated in future *Gaia* data releases with a longer baseline, which would increase the minimum a_{int} that we need to consider and thus reduce the maximum possible perturbation to \tilde{v} . Once this maximum is sufficiently small, it will be clear that WBs should remain in our sample despite receiving velocity perturbations from undetected CBs. This would correspond to a narrower range of a_{int} and most likely reduce f_{CB} .

These difficulties could explain why changing from a flat to a triangular logarithmic distribution of a_{int} reduces f_{CB} by 0.033 and narrowing the width from 1.5 dex to 1 dex reduces the inferred f_{CB} by 0.063. One can envisage that combining these changes would reduce f_{CB} by 0.1 from its nominal value of 0.63. Further fine-tuning of the a_{int} distribution could perhaps reduce f_{CB} to about 0.5. It seems unlikely that f_{CB} could be much smaller than this (see Appendix D). Even so, it is clear that unlike the inferred gravity law, the inferred incidence of CB companions is not as robust within its formal uncertainty. We therefore suggest that one should not read too much into the somewhat high f_{CB} inferred by our analysis.

5.1.4 The frequency of CB companions

Given that LOS contamination would cause a rising \tilde{v} distribution [equation (27)] and that \tilde{v} uncertainties are very small (Fig. 4), our model must explain the extended declining tail towards high \tilde{v} (Fig. 12) using mostly undetected CBs. This was also suggested by several other workers (Section 3.2) and expected *a priori* (BZ18). We argued previously that trends in the CB fraction with the WB separation are unlikely to mask the MOND signal (Fig. 16). However, this still leaves open the possibility that f_{CB} could have a different value to the 63 ± 1 per cent inferred by our analysis. Since the properties of the extended tail are to some extent degenerate with the inferred gravity law, we consider whether we can push this towards MOND by fixing $f_{\text{CB}} = 0.3$, which is close to half the somewhat high value inferred by our nominal analysis. The idea is that the model cannot easily fit the extended tail using CBs, so it might try to do so by changing the gravity law.

The results of this exercise are shown in Fig. 18. The inferred gravity law shifts moderately towards MOND, but it still remains much more consistent with the Newtonian prediction, which now faces just over 3σ tension. Since this reanalysis is the only one to substantially raise the inferred α_{grav} , we consider this model in more detail in Appendix D. We argue that since the likelihood of the best overall fit is lower than in our nominal analysis by a factor of $\exp(413)$, there is a 29σ preference for the latter. We trace this to a visible failure to match nearly all aspects of the WB data set when $f_{\text{CB}} = 0.3$ (Fig. D1).

5.2 Comparison with prior WBT results

Our approach to the WBT is very similar to PS23, who also considered the distribution of $(r_{\text{sky}}, \tilde{v})$. Our stellar mass estimates should be more reliable because we use a cubic rather than linear mass-luminosity relation and calibrate the resulting mass estimates with spectroscopic FLAME masses [equation (9)]. We calculate the relative velocity of each WB much more carefully, including an allowance for perspective effects due to the systemic RV. Our sample selection is also much stricter in several ways (Section 2.4), including

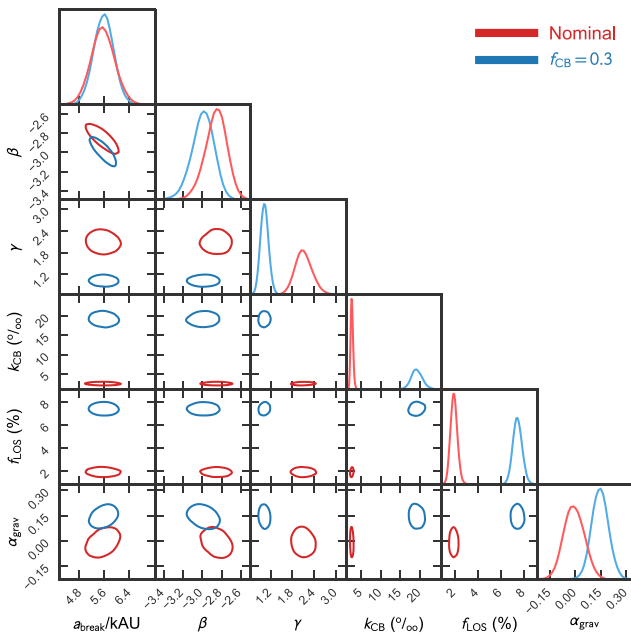


Figure 18. Triangle plot comparing the nominal analysis (red) to a revised analysis in which the likelihood of a star in our sample having an undetected CB companion is fixed at 30 per cent (Section 5.1.4).

through a quality cut on the carefully calculated \tilde{v} uncertainty. On the theoretical side, we improve upon their WB model by using orbits integrated in a rigorously calculated gravitational field (BZ18) and allow the eccentricity distribution to be superthermal, which seems very likely observationally (Hwang, Ting & Zakamska 2022). Our CB model is also an improvement because we allow both stars in a WB to have an undetected CB companion and model the CBs themselves in a somewhat different way, paying particular attention to the relevant range of CB orbit sizes and the inflation of the WB system mass due to undetected companion(s). Our treatment of chance alignments with unbound stars is simpler for computational reasons (Section 3.3), though it should still be adequate given that the inferred f_{LOS} is only about 2 per cent (Table 3). Our exploration of the parameter space is vastly more thorough thanks to an optimized algorithm (Banik, Pittordis & Sutherland 2021). In addition, we conduct a much broader range of extension analyses where we vary the sample selection and model assumptions, thus demonstrating the robustness of our results to a much greater extent (Section 5.1). Despite these changes, our overall conclusions are very similar with regards to how strongly Newtonian gravity is preferred over MOND (Section 4.1).

While this study was under review, two additional publications appeared which argue that the WBT prefers MOND (Chae 2023; Hernandez 2023). The study of Hernandez (2023) focused on building a very clean sample of WBs where CB contamination should be very small. While this could simplify the interpretation of the results, the sample size is unfortunately also rather small (see its fig. 8). In the following, we therefore focus on the work of Chae (2023). The particular issue that we identify could also affect the analysis of Hernandez (2023), though we have not checked this in detail.

The WB analysis of Chae (2023) uses the WB sample of El-Badry, Rix & Heintz (2021), but it attempts to infer the 3D separation and relative velocity from the available projected information instead of modelling into the space of the observables. Various complicated procedures are used to deal with the fact that WBs are expected to

have a wide range of eccentricities and to be viewed from a random direction at a random orbital phase. The eccentricity is constrained using the angle between the projected separation and relative velocity, making heavy use of the approach pioneered by Hwang, Ting & Zakamska (2022). The overall result is that MOND is preferred at a confidence of 10σ using the nominal sample of 26 615 WBs. This sample is not very well suited to the WBT due to a mass-dependent cutoff to the allowed \tilde{v} , but as this seems to have only a modest impact, we defer discussion of this issue to Appendix E.

Chae (2023) follow equation 4 of El-Badry, Rix & Heintz (2021) in neglecting the systemic RV and setting

$$v_{\text{sky}} = \bar{d}\mu_{\text{rel}}, \quad \mu_{\text{rel}} \equiv \begin{bmatrix} \mu_{\alpha,*}^{\text{rel}} \\ \mu_{\delta}^{\text{rel}} \end{bmatrix}, \quad (32)$$

where \bar{d} is the inverse variance weighted mean of the parallax distances to the stars A and B that make up the WB, $\mu_{\text{rel}} \equiv |\mu_{\text{rel}}|$ is the relative proper motion, $\mu_{\alpha,*}$ is the proper motion along the East-West direction, μ_{δ} is the proper motion along the North-South direction, and ‘rel’ superscripts denote relative values found by subtracting the value for star B from the value for star A (e.g. $\mu_{\delta}^{\text{rel}} \equiv \mu_{\delta}^{\text{A}} - \mu_{\delta}^{\text{B}}$). The proper motion $\mu_{\delta} \equiv \dot{\delta}$, where δ is the declination and an overdot denotes a time derivative. To account for the spherical geometry of quantities on the sky, the proper motion component $\mu_{\alpha,*} \equiv \dot{\alpha} \cos \delta$, where α is the right ascension. Both stars are assumed to be at the same distance, which should be correct on average if we assume that uncertainties in trigonometric distances are much larger than actual separations along the LOS. Neglecting the perspective effect due to the systemic RV should also be reasonable given the typically small angular separations of WBs (Section 2.3.1).

We now try to reproduce the claimed detection of MOND by Chae (2023). Without going into the details of the deprojection algorithm or the model for the CB and WB populations, a simple way to consider the WBT is to find the median \tilde{v} in different ranges of r_{sky}/r_M . Bearing in mind the above small caveat about perspective effects and the need to accurately estimate the stellar masses, these quantities can all be unambiguously calculated with no assumptions about how exactly WBs behave, how eccentric their orbits typically are, etc. The only assumption we need to make is that in the theories under consideration, WBs would very rarely have $\tilde{v} > 1.5$, so values much beyond this are indicative of some sort of contamination. It is therefore important to focus on only those WBs with $\tilde{v} \lesssim 2$ when calculating the median, since otherwise trends in the level of contamination could create the appearance of an acceleration-dependent trend in the WB dynamics (magenta line in Fig. 11). For the nominal sample of 26 615 WBs used by Chae (2023), the result of this exercise is shown in the left-hand panel of Fig. 19 based on masses and v_{sky} values from that study. Due to the larger sample size compared to our study, we are able to use 50 bins while keeping the scatter low. A rising trend is immediately apparent, especially for the subsamples with $\tilde{v} < 2$ or 2.5. This could arguably indicate that a limit of 1.5 is too restrictive and loses some of the MOND signal, which is possible given that the upper limit to \tilde{v} for a bound Milgromian WB is about 1.7 (BZ18). The results agree very well with the dashed grey line, which shows the shape that we expect in MOND (the normalization has been adjusted to fit the results at high accelerations or low r_{sky}/r_M). The median \tilde{v} undergoes some scatter but with little trend when $r_{\text{sky}}/r_M < 0.2$. It then rises rapidly at about the expected rate before flattening out. While we do not expect the observed rising trend at $r_{\text{sky}}/r_M \gtrsim 1.5$ (see Fig. 7), the median \tilde{v} is only rising quite slowly by this point and is broadly in agreement with the expected flat trend. All this seems to provide compelling evidence in favour of local WBs following Milgromian gravity.

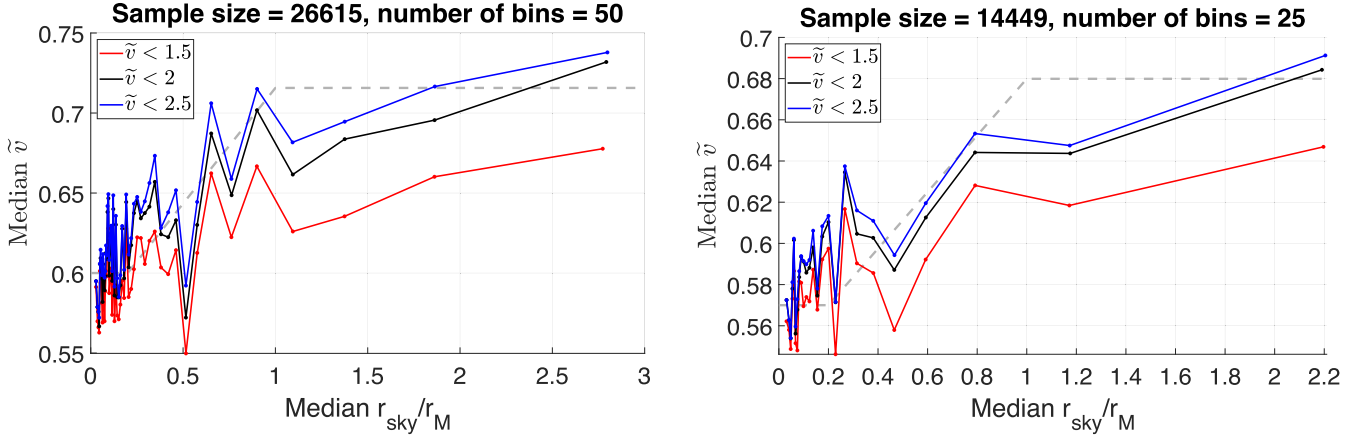


Figure 19. The median \tilde{v} as a function of the median r_{sky}/r_M for the nominal sample of Chae (2023), which requires both proper motion components of both stars in each WB to be more precise than 1 per cent (left-hand panel). This is tightened to 0.3 per cent in the right-hand panel, as done in their appendix B and Chae (2024). The sample size and number of bins are shown at the top. Similarly to Fig. 11, each point corresponds to the same number of WBs apart from the last point, which has a slightly different number due to rounding.

A critical consideration for the WBT is the accuracy of v_{sky} , which is the main source of uncertainty in \tilde{v} . We quantify its uncertainty very carefully in order to limit the \tilde{v} uncertainty of the WBs in our sample (Section 2.4.6). As argued there, since the main peak to the \tilde{v} distribution lies at $\tilde{v} \approx 0.5$ (Fig. 12) and \tilde{v} uncertainties would add in quadrature to the intrinsic width of the \tilde{v} distribution, we need the uncertainty in \tilde{v} to be $\lesssim 0.1$ in the regime most relevant to the WBT. An uncertainty of 0.1 would inflate the dispersion from 0.5 to $\sqrt{0.5^2 + 0.1^2} = 0.51$, implying only a 2 per cent broadening – too little to affect the inferred gravity law. While this approach might seem somewhat conservative, we have seen that it still leaves a reasonable sample size, partly because we impose a less strict limit to the \tilde{v} uncertainty outside the main peak region [equation (10)].

Unfortunately, Chae (2023) do not estimate the uncertainty in \tilde{v} or in v_{sky} . Their approach is to require a maximum 1 per cent uncertainty on the heliocentric proper motions $\mu_{\alpha,*}^A$, $\mu_{\alpha,*}^B$, μ_{δ}^A , and μ_{δ}^B of the stars in each WB. To see why this approach is not sufficient, suppose that both stars have a heliocentric velocity of 30 km s^{-1} in the East-West and North-South directions (the small difference due to WB orbital motion is not relevant here). The sky-projected heliocentric velocity of each star could then have an uncertainty of 300 m s^{-1} along each direction. Since we need to combine four proper motion components to calculate v_{sky} [equation (32)], this has an uncertainty of $\approx 300\sqrt{4}$ or 600 m s^{-1} . However, the Newtonian v_c of two Sun-like stars with a 10 kAU separation is only 420 m s^{-1} , implying a very poorly constrained \tilde{v} . This completely undermines the WBT because its main idea is that some typical measure of $\tilde{v} \propto \sqrt{g/g_N}$, making it absolutely essential to have precise constraints on \tilde{v} .

To try and address this issue, Chae (2023) conduct an extension analysis in their appendix B where the 1 per cent requirement mentioned above is tightened to 0.3 per cent. This is unlikely to be enough because in the previous argument, we would have a maximum allowed $\sigma(v_{\text{sky}})$ of 180 m s^{-1} rather than 600 m s^{-1} , where $\sigma(Q)$ is the uncertainty on any quantity Q . However, if the Newtonian v_c is 420 m s^{-1} and the mode of the \tilde{v} distribution is at 0.5, the intrinsic velocity dispersion is 210 m s^{-1} , which is not much larger than the upper limit to $\sigma(v_{\text{sky}})$. Alternatively, the plausible scenario described above would lead to $\sigma(\tilde{v}) = 180/420 = 0.43$, which is still too large for the WBT. Even so, we repeat the median \tilde{v} analysis with all four relevant heliocentric proper motions required to have an accuracy of at least 0.3 per cent. The sample size is almost halved

by this much stricter cut, so we use only half as many bins to reduce the scatter. Despite this, we still achieve a good range in r_{sky}/r_M and probe well into the asymptotic regime (right-hand panel of Fig. 19). A rising trend is evident here, though the agreement with the MOND expectation is somewhat less good. Importantly, it is not really possible to fit the results with a flat line, as would be expected in Newtonian gravity.

A more rigorous but still very simple way to estimate the uncertainty in v_{sky} is provided by equation 5 of El-Badry, Rix & Heintz (2021). If we neglect other sources of error like perspective effects and the small error in \bar{d} , we can approximate that the only uncertain term is μ_{rel} . Its uncertainty is

$$\sigma(\mu_{\text{rel}}) = \frac{\sqrt{\sigma^2(\mu_{\alpha,*}^{\text{rel}}) + \sigma^2(\mu_{\delta}^{\text{rel}})}}{\mu_{\text{rel}}}, \quad (33)$$

$$\sigma(\mu_{\alpha,*}^{\text{rel}}) = \sqrt{\sigma^2(\mu_{\alpha,*}^A) + \sigma^2(\mu_{\alpha,*}^B)}, \quad (34)$$

$$\sigma(\mu_{\delta}^{\text{rel}}) = \sqrt{\sigma^2(\mu_{\delta}^A) + \sigma^2(\mu_{\delta}^B)}. \quad (35)$$

This result is based on the fact that at first order in $\sigma(\mu_{\text{rel}})/\mu_{\text{rel}}$, the uncertainty in μ_{rel} arises only from the proper motion uncertainty along the direction $(\mu_{\alpha,*}^{\text{rel}}, \mu_{\delta}^{\text{rel}})$ because uncertainty in μ_{rel} in the orthogonal direction mainly affects its direction rather than its magnitude. This approximation becomes inaccurate if $\sigma(\mu_{\text{rel}}) \gtrsim \mu_{\text{rel}}$, so a Monte Carlo approach might be preferable even if we could assume that the uncertainty arises entirely from uncorrelated astrometric errors in the proper motions.

With the above approximations, the uncertainty in v_{sky} is simply

$$\sigma(v_{\text{sky}}) = \bar{d}\sigma(\mu_{\text{rel}}). \quad (36)$$

Dividing this by the Newtonian v_c then yields an estimated \tilde{v} uncertainty. We suggest that such an approach would capture the major uncertainties affecting \tilde{v} because the mass and systemic distance should both have small uncertainties, while perspective effects should likewise have little effect (Section 2.3.1).

We can now check whether the result of Chae (2023) is robust against a quality cut based on the estimated \tilde{v} uncertainty of each WB. We begin by quantifying the proportion of WBs in each r_{sky}/r_M bin with $\sigma(\tilde{v}) > 0.1 \max(1, \tilde{v}/2)$, the cut imposed in our own analysis [equation (10)]. The result is shown in Fig. 20. It is clear that most WBs in the asymptotic regime fail to pass our quality cut on $\sigma(\tilde{v})$.

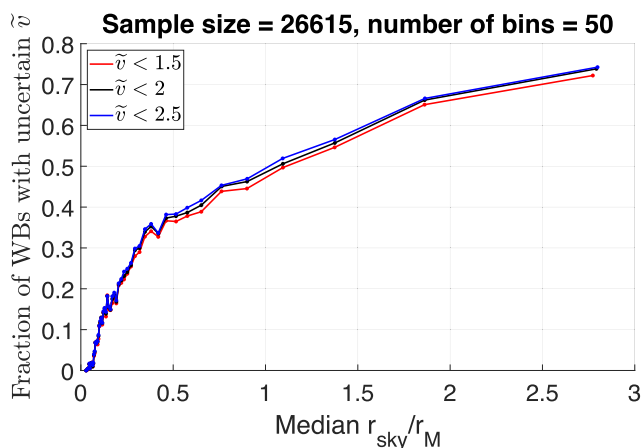


Figure 20. The fraction of WBs in each r_{sky}/r_M bin of the nominal Chae (2023) sample with $\sigma(\tilde{v}) > 0.1 \max(1, \tilde{v}/2)$. This applies to most of the WBs in the asymptotic regime critical to testing MOND.

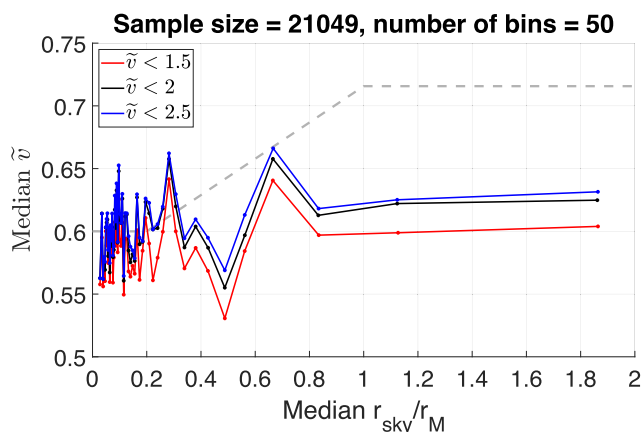


Figure 21. Similar to the left-hand panel of Fig. 19, but with the Chae (2023) nominal sample restricted to WBs with an estimated $\sigma(\tilde{v}) < 0.1 \max(1, \tilde{v}/2)$ according to equations (33) and 36. The median \tilde{v} is now flat with respect to our proxy for the WB acceleration thanks to this cut on the estimated \tilde{v} uncertainty, which is critical to reliably conducting the WBT.

This raises the possibility that the MOND-like trend in Fig. 19 is caused by measurement errors broadening the \tilde{v} distribution at low accelerations.

To analyse the data set of Chae (2023) more robustly, we restrict their nominal sample to the same limiting $\sigma(\tilde{v})$ as we used in our own analysis [equation (10)]. The result is shown in Fig. 21. The sample size only decreases by about 1/5, allowing us to continue using 50 bins in r_{sky}/r_M , albeit with a somewhat reduced range given that the sample is reduced by half at the MOND radius and slightly more beyond that. Even so, we still have many WBs well into the asymptotic regime. These show that the median \tilde{v} is now almost completely flat with respect to r_{sky}/r_M . There is clearly a major disagreement with the dashed grey line showing the MOND expectation calibrated to the high-acceleration end. This dramatic difference is caused by removing only 1/5 of the WBs, indicating that the issue we identified is indeed extremely serious.

The MOND-like rising median \tilde{v} trend in the nominal sample of Chae (2023) persists despite an apparently tight 0.3 per cent cut on the proper motion errors, which leads to an even smaller sample size than in Fig. 21. This highlights that the important quantity for the

WBT is the relative velocity of the stars in each WB rather than the heliocentric velocity of each star. While one can envisage that a sufficiently restrictive cut on the latter would be sufficient, this is simply not targeting the relevant quantity, so the sample size is likely to become insufficient before the cut has been adequately tightened for the purpose of the WBT. As an exercise in this direction, we found that requiring a 0.1 per cent precision reduces the sample to only 4977 WBs and makes the median \tilde{v} nearly flat with respect to r_{sky}/r_M , so perhaps this cut is tight enough. However, the sample size can be much larger while still maintaining sufficient accuracy on \tilde{v} provided that the quality cut is targeted at \tilde{v} , or at least at μ_{rel} .

The importance of this issue stems from the fact that for the same $\sigma(v_{\text{sky}})$ in m s^{-1} , $\sigma(\tilde{v}) \propto \sqrt{r_{\text{sky}}}$ because the Newtonian $v_c \propto 1/\sqrt{r_{\text{sky}}}$, making it increasingly difficult to pass a strict quality cut on $\sigma(\tilde{v})$ (Fig. 20). Without such a cut, as we get to large r_{sky} and thus also r_{sky}/r_M , the \tilde{v} distribution might be broadened by measurement errors to a greater extent, pushing up the median \tilde{v} and creating a fake Milgromian signal. This raises the more general issue that any systematic trend with r_{sky} can appear to be an acceleration-dependent trend. In principle, one could disentangle the two because our proxy for the WB internal acceleration depends on both r_{sky} and M . However, the former is by far the main factor because of its large range and the fact that r_M scales only as \sqrt{M} , making the WB mass relatively less important. There is also an issue with the data set of Chae (2023) pertaining to the masses that makes it difficult to use a broad range (Appendix E). One therefore has to be cautious before concluding that any trend discovered in the WB data is best understood as causally related to their internal acceleration.

Given our results in this section, it seems extremely unlikely that local WBs prefer Milgromian over Newtonian gravity once the WB sample is selected carefully to ensure that \tilde{v} is accurately known. It should be straightforward to alter the quality cut related to the astrometric accuracy by focusing on how precisely we know the relative velocity within each WB, since this is what enters into the calculation of the physically relevant \tilde{v} parameter.¹⁶ We stress that even if \tilde{v} is not calculated explicitly, the WBT is ultimately about the ratio between the relative velocity and the Newtonian circular orbit prediction because the WB orbits are too long to directly measure the acceleration (Banik & Kroupa 2019a). The uncertainty on this velocity ratio is therefore of critical importance to the WBT and must be quantified, with systems rejected if the uncertainty is too large.

This concern likely also applies to the analysis of Hernandez (2023): its section 4 states that v_{sky} should exceed $1.5 \times$ its uncertainty. This translates to a fractional \tilde{v} accuracy of 2/3, so a system with $\tilde{v} = 1$ might have $\sigma(\tilde{v}) = 0.6$ (this concern also applies to Hernandez et al. 2024). Addressing this issue similarly to our study would reduce the already small sample size. However, our work suggests that the loss might be modest, especially given the high overall quality of the WB sample used in their study.

5.3 The MOND interpolating function

Our results for the WBT depend on the assumed MOND interpolating function because local WBs are subject to the Galactic external field of magnitude $1.8 a_0$ (Section 3.1; see also section 7.1 of BZ18). Thus, our falsification of MOND could in principle be avoided with a sufficiently rapid transition to Newtonian behaviour when $g_N > a_0$.

¹⁶This would allow to relax the requirement for the heliocentric proper motion components to have a 1 per cent precision, increasing the sample size somewhat.

Table 4. The enhancement factor η to the radial gravity [equation (15)] and our corresponding estimate for α_{grav} [equation (39)] with different MOND formulations and interpolating functions. Results differ little between AQUAL and QUMOND and between the MLS and simple functions.

Interpolating function	AQUAL		QUMOND	
	η	α_{grav}	η	α_{grav}
Simple	1.4056	0.96	1.4228	1
MLS	1.3508	0.84	1.3692	0.88
Standard	1.0661	0.17	1.0726	0.18
Sharp	1	0	1	0

However, this is in strong tension with rotation curve constraints. This is because the enhancement to g_N needed to fit the host galaxy rotation curve is very similar to the predicted enhancement to the gravity binding a WB in the asymptotic regime. This remains true in both AQUAL and QUMOND, which give almost the same boost for the same interpolating function (Fig. 7). We clarify this by briefly giving the main analytic results relevant to EFE-dominated WBs and thereby estimate the predicted α_{grav} with different MOND formulations and interpolating functions (for further details, see sections 2.2 and 7.1 of BZ18).

The WBT is mostly sensitive to η [equation (15)] in the asymptotic regime. In QUMOND, equation 16 of BZ18 tells us that

$$\eta = v_e \left(1 + \frac{K_e}{3} \right), \quad K_e \equiv \frac{\partial \ln v_e}{\partial \ln g_{N,e}}, \quad (37)$$

where v_e is the QUMOND interpolating function if we set its argument to the Newtonian-equivalent Galactic gravity $g_{N,e}$ [equation (14)]. This result matches equation 57 of Milgrom (2010).

AQUAL does not directly use v but instead uses the interpolating function μ , which is defined so that $\mu g = g_N$ in spherical symmetry. We can talk of AQUAL and QUMOND theories as having the ‘same’ interpolating function if they give the same relation between g and g_N in spherical symmetry, which requires that $\mu(g)v(g_N) = 1$. According to equation 35 of BZ18, the AQUAL version of equation (37) is

$$\eta = \frac{\tan^{-1} \sqrt{L_e}}{\mu_e \sqrt{L_e}}, \quad L_e \equiv \frac{\partial \ln \mu_e}{\partial \ln g_e}, \quad (38)$$

where L_e is the AQUAL interpolating function considering only the Galactic gravity g_e . We can find L_e from the QUMOND K_e using the relation $(1 + L_e)(1 + K_e) = 1$ (equation 38 of BZ18). The angle-averaged enhancement to g_N shown above matches that stated in equation 65 of Milgrom (2010) once we use the relation $\sin^2 \theta = \tan^2 \theta / (\tan^2 \theta + 1)$.

The observable in the WBT is the relative velocity and this is related to the circular velocity, which scales as $\sqrt{\eta}$. Moreover, the parameter α_{grav} only captures deviations from the Newtonian result because it is defined to be 0 in the Newtonian case, when $\eta \equiv 1$. Thus, we can approximate that the predicted $\alpha_{\text{grav}} \propto \sqrt{\eta} - 1$. Since the normalization must be chosen such that $\alpha_{\text{grav}} = 1$ if we use QUMOND with the simple interpolating function,

$$\alpha_{\text{grav}} = \frac{\sqrt{\eta} - 1}{0.193}. \quad (39)$$

Table 4 shows η and the estimated α_{grav} for different MOND formulations and interpolating functions. Differences between AQUAL and QUMOND are very small, as stated in section 5.1 of Milgrom (2010) and explained further in section 7.2 of BZ18, which is devoted to a comparison of the two MOND formulations. The simple and MLS ($v^{-1} = 1 - \exp(-\sqrt{g_N/a_0})$; McGaugh 2008; Famaey &

McGaugh 2012; McGaugh, Lelli & Schombert 2016) interpolating functions give similar predictions for the WBT because they are numerically very similar. The main difference is that the MLS function converges to Newtonian gravity much faster when $g \gtrsim 10 a_0$ due to the exponential term, but this is not important for the WBT because WB observations are sensitive to the gravity binding it rather than merely how much that deviates from g_N .¹⁷ Compared to the simple and MLS cases, the standard interpolating function significantly reduces the predicted signal for the WBT, though even then the predicted $\alpha_{\text{grav}} = 0.18$ is still much higher than inferred by our nominal MCMC analysis (bottom right panel of Fig. 14). But since some variations to the modelling assumptions do increase the inferred α_{grav} somewhat (Table 3), it can be argued that the WBT is marginally consistent with the standard interpolating function

$$v^2 = \frac{1}{2} + \sqrt{\frac{1}{4} + \left(\frac{a_0}{g_N}\right)^2}. \quad (40)$$

However, this is in tension with rotation curve constraints (section 7.1 of BZ18). We revisit their arguments in more detail below.

Equations (37) and (38) show that $\eta \approx v_e = 1/\mu_e$ up to a modest correction for azimuthal averaging. This is because v_e can be arbitrarily large, but K_e is always between $-1/2$ and 0 while L_e is between 1 and 2, with low K_e and high L_e corresponding to the deep-MOND limit ($K_e = -0.26$ in the Solar neighbourhood). As a result, the enhancement to the orbital velocities of WBs in the asymptotic regime should be only slightly less than the enhancement to the Newtonian baryonic rotation curve of the parent galaxy in the vicinity of the WBs. This means a very stringent null detection of MOND effects in the WBT can only be reconciled with MOND if the Galactic rotation curve at the Solar circle has almost the same amplitude as it would do in Newtonian gravity with baryons alone.

We test this in Fig. 22, where the cyan squares with error bars show binned observational results and the dotted blue vertical line shows the Solar circle radius of 8.2 kpc (the uncertainty on this is very small; Gravity Collaboration 2019). The largest uncertainty in the kinematics comes from the peculiar (non-circular) velocity of the Sun and hence on the actual value of $v_{e,\odot}$, which normalizes the entire curve. However, the resulting uncertainty has been substantially mitigated by the direct measurement of the Solar System’s acceleration relative to distant quasars (Gaia Collaboration 2021b), which is a direct probe of g_e . Besides the kinematics, there is also some systematic uncertainty in the Newtonian rotation curve of the Galactic disc because we do not perfectly know its actual mass-to-light ratio and gas content. With these small caveats in mind, the Newtonian rotation curve of the baryons alone (orange line) is clearly below the actual rotation curve at the Solar circle, which is also much flatter. Since scaling the Newtonian rotation curve (e.g. to raise the mass-to-light ratio) would make the predicted decline even steeper, it is clear that MOND can only work with an interpolating function that significantly enhances Newtonian gravity at the Solar circle. Indeed, the solid green line shows that QUMOND with the simple interpolating function fits rather well – even the bumps and wiggles apparent in the observations are reproduced nicely if we consider similar features in the Galactic surface density profile (McGaugh 2018). However, the required significant Milgromian enhancement to the local amplitude of the rotation curve contradicts

¹⁷The simple interpolating function provides a good match to the velocity dispersion profiles of elliptical galaxies up to almost $100 a_0$ (Chae & Gong 2015; Chae et al. 2019, 2020a), beyond which the MOND enhancement to g_N must be very small.

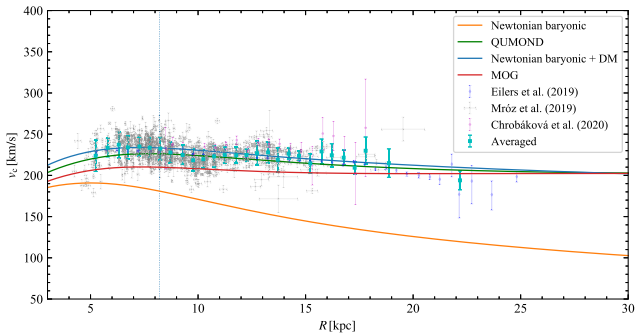


Figure 22. The MW rotation curve as measured by averaging different observational determinations (cyan squares with error bars; Eilers et al. 2019; Mróz et al. 2019; Chrobáková et al. 2020) and as predicted in different gravity theories using the baryonic mass distribution from Wang, Hammer & Yang (2022). Notice that the Newtonian model with only baryons (orange line) falls well short of the observed rotation curve at the Solar circle radius of 8.2 kpc (dotted blue vertical line). This deficiency can be rectified by adding dark matter (blue line), but the halo parameters must be inferred from the data and are not predicted *a priori*. An alternative solution with very little flexibility is QUMOND with the simple interpolating function [green line; equation (13)], but this only works because this function predicts a significant enhancement to Newtonian gravity at the relevant acceleration (compare the simple and sharp interpolating functions in Table 4). This enhancement should also be detectable in local WBs. Reproduced from fig. 1 of Zhu et al. (2023), modified by Haixia Ma.

our stringent null detection of MOND effects in the WBT, which implies an enhancement by at most a few percent. This contradiction is underpinned by the directly measured acceleration of the Solar System relative to distant quasars (Gaia Collaboration 2021b), which precludes substantial deviations from the kinematically deduced g_e (Section 3.1). If we very conservatively assume that $v_{c,\odot}$ is at most 10 km s^{-1} higher than our adopted 232.8 km s^{-1} , then the stronger EFE would only reduce the QUMOND α_{grav} for the simple interpolating function from 1 to 0.92, while an even higher $v_{c,\odot} = 250 \text{ km s}^{-1}$ would give $\alpha_{\text{grav}} = 0.87$. Such a large deviation of $v_{c,\odot}$ from our adopted value is almost inconceivable given the many decades of research on this issue. Turning instead to the possibility of a much sharper interpolating function, the required enhancement to the baryonic surface density would need to be truly substantial, which seems rather unlikely.

Another interesting aspect of Fig. 22 is the failure of Moffat gravity (MOG; Moffat 2006), whose prediction is shown as the red line. MOG modifies gravity only beyond a certain distance, so it passes the WBT (see section 2.3 of Roshan et al. 2021a). However, MOG underpredicts the Galactic rotation curve (as shown previously by Negrelli et al. 2018). Moreover, a joint fit to the velocity dispersion profile and star formation history of Dragonfly 44 rules out MOG at 5.5σ confidence (Haghi et al. 2019). This shows how difficult it is for a modified gravity theory to remain consistent with the WBT and simultaneously explain galaxy dynamics, even if we neglect extragalactic data.

Of course, one should also consider the rather precise constraints on MOND available nowadays from extragalactic rotation curves. In particular, the Spitzer Photometry and Accurate Rotation Curves (SPARC; Lelli, McGaugh & Schombert 2016) catalogue reveals a tight RAR, which we show in Fig. 23 as the ratio between the observed g at fixed g_N and the prediction g_v with different interpo-

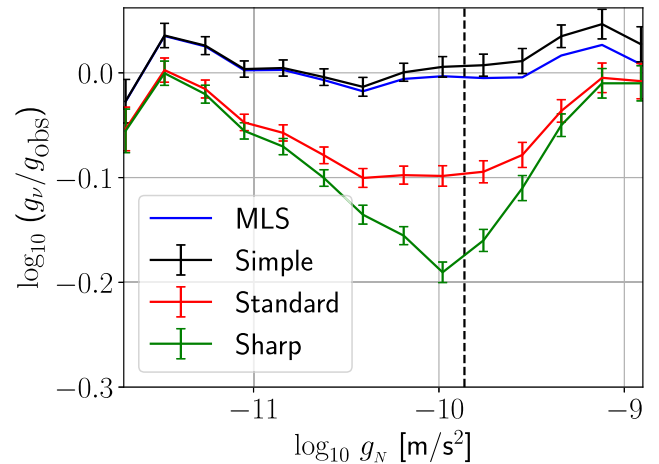


Figure 23. The ratio between the observed g at fixed baryonic g_N (McGaugh, Lelli & Schombert 2016) and the prediction g_v according to different MOND interpolating functions for the canonical $a_0 = 1.2 \times 10^{-10} \text{ m s}^{-2}$ (Begeman, Broeils & Sanders 1991; Gentile, Famaey & de Blok 2011; Lelli et al. 2017b). Observational uncertainties are found by dividing the dispersion in $\log_{10} g$ at each g_N by \sqrt{N} , where N is the number of data points in the bin. Due to the numerical similarity of the simple and MLS functions (black and blue lines, respectively); error bars are omitted for the latter, but they are shown on the standard interpolating function [red; equation (40)] and for the case of an infinitely sharp transition between the Newtonian and MOND regimes [green; equation (41)]. The dashed vertical line shows the Solar circle g_N assuming the simple interpolating function adopted for this study [equation (13)].

lating functions.¹⁸ The error bars show the logarithmic dispersion in g_N about a smooth relation scaled down by \sqrt{N} , where N is the number of data points in the g_N bin. It is clear that the MLS and simple functions are numerically rather similar and both match the observations fairly well. However, the standard interpolating function [equation (40)] deviates very substantially from observations (this is also apparent in fig. 1 of Tian & Ko 2019). As with other analyses, the crucial assumption is the stellar mass-to-light ratio from stellar population synthesis models (Schombert, McGaugh & Lelli 2022), which would need substantial modifications to allow the standard interpolating function to match the data.

This discrepancy is worsened by using an infinitely sharp transition between the Newtonian and deep-MOND regimes, i.e.

$$v = \max\left(1, \sqrt{\frac{a_0}{g_N}}\right). \quad (41)$$

This yields no enhancement to gravity at the Solar circle ($\alpha_{\text{grav}} = 0$), so the results of the WBT work best with this function given that our posterior inference on $\alpha_{\text{grav}} = -0.021^{+0.065}_{-0.045}$. Even the modest enhancement predicted by the standard interpolating function causes 3σ tension with our WB results. At the same time, this enhancement is too little to explain extragalactic rotation curves across almost the full range of g_N that they probe (Fig. 23). Similar conclusions can be drawn from the Galactic rotation curve, which barring major systematics shows that $g > g_N$ by about 50 per cent at the Solar circle even though both slightly exceed a_0 (Fig. 22). Therefore, it seems impossible for MOND to simultaneously match galaxy dynamics

¹⁸A unique relation is not expected in modified gravity theories because disc galaxies are not spherically symmetric, but deviations from the spherically symmetric relation should be very small (Jones-Smith et al. 2018; Chae 2022).

and *Gaia* data on local WBs. This falsifies classical modified gravity versions of MOND at high significance.

5.4 Broader implications

So far, we have assumed that MOND should be understood as a modification to Newtonian gravity in the weak-field regime [equation (1)]. Given the failure of this approach in local WBs, it is helpful to consider MOND as a modification to inertia at low accelerations while preserving Newtonian gravity (Milgrom 1994, 2011, 2022). This may avoid significant MOND effects in local WBs because their internal orbital motion has a much higher frequency than their Galactocentric orbit (e.g. see fig. 14 of BZ18). We can consider this in terms of an effective Newtonian external field $g_{N,e}^{\text{eff}}$, defined such that the boost to the gravity binding a WB is still given by equation (37), but with argument $g_{N,e}^{\text{eff}}$ instead of the Newtonian gravity $g_{N,e}$ sourced by the rest of the Galaxy. Since the MOND model considered in the main part of our study predicts a 42 per cent enhancement to the average radial gravity (Table 4) and our results show that $\alpha_{\text{grav}} < 0.2$ at quite high confidence (Table 3), we can interpret the WBT as constraining the average gravity binding a WB to within 8 per cent of the Newtonian prediction in the asymptotic regime where the Galactic gravity dominates over the internal gravity. The corresponding lower limit on $g_{N,e}^{\text{eff}}$ is $7.8 a_0$ ($4.6 a_0$) for the simple (MLS) interpolating function. In both cases, this is several times larger than the expected value of $\approx 1.2 a_0$ (Section 3.1.1). The high $g_{N,e}^{\text{eff}}$ could indicate that the Galactic EFE suppresses the gravity of a local WB by much more than we have assumed, making it very nearly Newtonian.

Modified inertia theories of MOND are still at an early stage of development, so it is not yet possible to conduct detailed simulations of situations like interacting galaxies with a low degree of symmetry. Even so, focusing on systems where the predictions are clear led to the first attempt to distinguish whether MOND is better understood as a modification to gravity or inertia (Petersen & Lelli 2020). The results mildly prefer modified gravity at $\approx 1.5\sigma$ confidence. Modified inertia theories make the very strong prediction that test particles on circular orbits will have a unique relation between their kinematic acceleration and the gravity g_N that they experience. As a result, the inner and outer parts of galaxy rotation curves should fall on the same RAR. A difference is expected in modified gravity, essentially because the vertical gravity just outside the disc plane is higher nearer the galactic centre, reducing the MOND enhancement (Banik, Milgrom & Zhao 2018). Recently, a stacked analysis of SPARC rotation curves identified a 6.9σ difference between the RAR traced by data points from the inner and outer parts of galaxy rotation curves if the observational uncertainties are taken at face value (Chae 2022). The difference in the RAR evident in their fig. 6 is similar to that expected in modified gravity formulations of MOND. This could be a coincidence caused by unknown systematic errors, but this seems unlikely because almost the same signal is detected when excluding galaxies with a significant bulge or with a high luminosity, which could be signs of a greater degree of pressure support and thus possibly mean a larger uncertainty on the data points at low radii. This suggests that we should take seriously the observed difference in the RAR traced by the inner and outer parts of rotation curves. The high formal significance of the difference evident in their fig. 7 is *a priori* not expected in the modified inertia interpretation of MOND. One caveat is that gas motions are not perfectly circular, so until we have a fully fledged formulation of modified inertia theories capable of handling somewhat eccentric orbits, it is impossible to fully conclude on whether the results of Chae (2022) indeed rule out

such theories. But taking their results at face value and given also that the WBT rules out MOND as modified gravity (though see Milgrom 2023a, b), we most likely need to fundamentally change MOND if it is to survive the latest constraints, which questions its validity on all scales.

WBs are not the only bound systems that challenge MOND. Its predicted gravity in galaxy clusters generally falls short of the observed value, though one can assume a non-baryonic dark mass component that makes up about half of the gravitating mass (e.g. Sanders 2003; Angus, Famaey & Diaferio 2010). This discrepancy was recently illustrated in terms of the cluster RAR, as shown in fig. 8 of Eckert et al. (2022a) and fig. 5 of Li et al. (2023). The RAR traced by galaxy clusters lies about 0.3 dex ($2\times$) above the RAR traced by galaxies when $g_N \gtrsim 0.1 a_0$ (McGaugh 2020, and references therein). This gap narrows further and almost vanishes at the low acceleration end. The discrepancy with the Newtonian expectation without DM is of course larger at about 0.8 dex or $6\times$, which matches the ratio between total and baryonic mass in Λ CDM fits to the power spectrum of anisotropies in the cosmic microwave background radiation (Planck Collaboration VI 2020).

The WBT tests MOND on a much smaller scale than the traditional galaxy-scale tests with rotation curves, velocity dispersions, and weak lensing (section 3 of Banik & Zhao 2022, and references therein). This raises the possibility of suppressing MOND effects at short range, for instance with a length-dependent cutoff to any MOND enhancement to gravity below a scale of 0.1 pc. Babichev, Deffayet & Esposito-Farèse (2011) proposed an extended version of the Vainshtein screening mechanism in massive gravity (Vainshtein 1972; Babichev, Deffayet & Ziour 2010) in which MOND effects are suppressed below a length $r_B \propto M^{1/4}$ around a point mass M . Since r_B rises slower than the MOND radius ($r_M \propto M^{1/2}$), we get that in a sufficiently massive system, the suppression of MOND effects arises only in the Newtonian regime and thus has no effect. But in a very low mass system, a new regime appears where the distance r lies in the range $r_M < r < r_B$. In this regime, MOND as classically formulated predicts a departure from Newtonian gravity, but the behaviour becomes Newtonian with the extended screening mechanism (Babichev, Deffayet & Esposito-Farèse 2011).

Another way to think of this is in terms of the phantom dark matter (PDM) density ρ_{pdm} generated by MOND, which is defined such that

$$\nabla \cdot \mathbf{g} \equiv -4\pi G (\rho + \rho_{\text{pdm}}), \quad (42)$$

where \mathbf{g} must be found by solving equation (12) and ρ is the physical mass density. Since $g \propto \sqrt{M}/r$ in the deep-MOND limit [equation (1)], $\rho_{\text{pdm}} \propto \sqrt{M}/r^2$. If we focus on the PDM density at $r = r_B \propto M^{1/4}$, we get that ρ_{pdm} is a constant. Thus, we can think of the model as providing a maximum limit to $|\rho_{\text{pdm}}|$ (the modulus is needed because the PDM density can be negative in more complicated geometries; see Milgrom 1986b; Oria et al. 2021). Such an upper limit arises in some attempts to unify the acceleration discrepancies in galaxies with the late-time accelerated expansion of the Universe (Zhao 2007). We can use the WBT to estimate an upper bound on ρ_0 , the maximum possible $|\rho_{\text{pdm}}|$. Since our analysis indicates that $\alpha_{\text{grav}} < 0.2$ at rather high confidence (Table 3), we can assume that there is at most a 4 per cent enhancement to the orbital velocities of WBs in the asymptotic regime. This corresponds to an enclosed phantom mass equal to 8 per cent of the actual mass in the stars. Assuming our results are sensitive down to $M = 1 M_\odot$ and up to separations of $r = 20$ kAU or 0.1 pc, we can estimate that

$$\rho_0 < 0.08 M_\odot \div \frac{4\pi r^3}{3} = 20 M_\odot/\text{pc}^3. \quad (43)$$

This is three orders of magnitude above the local DM density inferred from a Newtonian dynamical analysis of Galactic data (Read 2014; Hagen & Helmi 2018; Salomon et al. 2020; de Salas & Widmark 2021). While we expect ρ_{pdm} to be somewhat higher in a dwarf galaxy due to its lower mass, it is clear that the application of MOND to galaxies would be unaffected by a maximum limit to $|\rho_{\text{pdm}}|$ of order $20 M_{\odot}/\text{pc}^3$. The impact on larger scales would be even less significant, leaving open the issue of how MOND might be reconciled with observations of galaxy clusters and larger structures.

A maximum limit to $|\rho_{\text{pdm}}|$ capable of reconciling MOND with the WBT would substantially reduce MOND effects in the Solar System, invalidating attempts to explain some peculiarities of the Kuiper Belt using MOND (Brown & Mathur 2023; Migaszewski 2023) but improving the agreement with the rather tight bound on non-standard effects provided by Cassini radio tracking measurements of the Earth-Saturn range (Hees et al. 2014, 2016). This bound is already so tight that it formally rules out even the exponentially truncated MLS interpolating function at 8σ confidence (see equation C15 of Brown & Mathur 2023), so it could be difficult to reconcile Solar System ephemerides with galaxy dynamics in classical modified gravity theories of MOND (Desmond et al., in preparation). Since the PDM density at the MOND radius of an isolated point mass M scales as $\rho_{\text{pdm}} \propto 1/\sqrt{M}$ and we are roughly at the MOND radius of the MW with $M = 7 \times 10^{10} M_{\odot}$ (Banik & Zhao 2018b), adding ρ_0 as a new constant of nature would suppress MOND effects only in systems up to six orders of magnitude less massive than the MW, i.e. up to about $M = 10^5 M_{\odot}$. This is in the regime of massive star clusters and globular clusters, which might help to explain why the internal kinematics of the outer halo globular cluster NGC 2419 are consistent with Newtonian expectations despite feeling little gravity from the Galaxy by virtue of its distance (Ibata et al. 2011a, b). While it has been argued that the observations are consistent with MOND as classically understood (Sanders 2012a, b), this requires one to invoke observational systematics like a radially varying polytropic equation of state. Since the mass of NGC 2419 is $9 \times 10^5 M_{\odot}$, a clear detection or exclusion of MOND effects here would help to constrain how MOND effects must be suppressed on small scales to ensure consistency with the WBT. Interestingly, MOND correctly predicts the velocity dispersions of isolated LG dwarf galaxies down to $10^6 M_{\odot}$ (fig. 3 of McGaugh et al. 2021), while tidal stability considerations of Fornax Cluster dwarfs reach down to about the same mass (fig. 7 of Asencio et al. 2022). It is therefore possible that MOND as classically formulated breaks down at lower masses and that hints of this are already apparent in NGC 2419. A problem with this scenario is the asymmetric tidal tails of star clusters in the Solar neighbourhood, which have been argued to favour MOND (Kroupa et al. 2022; Pflamm-Altenburg et al. 2023). Further investigation is needed to see if the results can be explained in Newtonian gravity with a more complicated Galactic model that includes bars and spiral arms, though recent work indicates that the bar is not sufficient by itself for any pattern speed (Thomas et al. 2023). If MOND effects persist down to order $10^3 M_{\odot}$ but are not apparent in the WBT, then this would give a tight constraint on where classical MOND breaks down.

A limit to the PDM density might also limit the predicted Milgromian enhancement to the vertical gravity of the Galactic disc, which for now is neither confirmed nor rejected (Zhu et al. 2023). If it becomes clear that the enhancement is smaller than predicted, this would be interesting because the Galactic rotation curve works very well in MOND (McGaugh 2018; Zhu et al. 2023) and probes scales which are not much larger than the disc thickness, thereby

pinning down the boundary between where MOND works and where additional physics must be invoked to suppress MOND effects. While such tests would be useful, for MOND to be valid on kpc scales and ultimately provide the correct explanation for the RAR, we have to invoke at least two new fundamental constants (a_0 and ρ_0 or a fundamental length). This seems rather contrived and not in line with Occam’s Razor, especially as ρ_0 is not obviously related to any cosmological density scale (unlike a_0) and cannot be substantially reduced below our estimate in equation (43) if we are to preserve the successes of MOND in dwarf galaxies (McGaugh et al. 2021; Asencio et al. 2022). We therefore discuss other theories that may be able to explain galaxy dynamics while passing the WBT.

Since we do not have a fully fledged quantum gravity theory, the idea of emergent gravity (EG; Verlinde 2017) is that gravity is an entropic force arising from underlying microscopic degrees of freedom that are currently not understood – and do not always need to be, much like how a very detailed understanding of molecules is not needed to understand the ideal gas law. EG faces major issues on galaxy cluster scales (Tamosiunas et al. 2019) and fails in galaxies because it predicts hook-shaped deviations from a single universal RAR that cannot be hidden within the uncertainties (Lelli, McGaugh & Schombert 2017a). While its prediction for the WBT is unclear, EG as presently formulated is strongly inconsistent with Solar System ephemerides (Hees, Famaey & Bertone 2017; Chan & Lee 2023). These precise constraints also rule out gravitational dipoles (Hajdukovic 2020; Banik & Kroupa 2020a) and scale-invariant dynamics (Maeder & Gueorguiev 2020; Banik & Kroupa 2020b). The WBT greatly exacerbates the already severe difficulties faced by modified gravity theories in explaining the RAR without DM while leaving no discernible trace in highly precise Solar System ephemerides, which our work effectively extends to 20 kAU at lower precision. We thus explore whether some form of DM on galaxy scales might provide a tight RAR (Famaey & McGaugh 2012) while lessening the very severe difficulties encountered by the standard CDM approach in other respects (Kroupa et al. 2010; Kroupa 2012, 2015; Di Valentino 2021; Banik & Zhao 2022, and references therein). We stress that the failures of Λ CDM identified in those works are unrelated to the validity of any alternatives proposed as solutions.

A hybrid MOND-DM model is provided by superfluid dark matter (SFDM; Berezhiani & Houry 2015). Its basic idea is that galaxies are embedded in DM haloes, but their total mass within the region traced by rotation curves is not significant compared to the baryonic mass. Instead, the flat rotation curve problem is solved by postulating additional non-gravitational interactions between the baryons. These are mediated by phonons propagating in the DM halo, which is possible in its central superfluid core. On larger scales, SFDM reduces to Λ CDM as the superfluid phase only arises at low temperature and high density, so it arises in galaxies but not in galaxy clusters. SFDM should alleviate the problem faced by Λ CDM with the fast observed rotation speeds of galaxy bars (Roshan et al. 2021b) thanks to reduced dynamical friction on subsonic flows like a rotating galaxy bar (Berezhiani, Elder & Houry 2019). However, the LG satellite planes are still very difficult to understand in SFDM because their anisotropy strongly suggests a tidal origin but their internal velocity dispersions imply some enhancement to the forces binding the satellites, which is hard to understand as the more distant satellites in these structures would be outside the superfluid portion of the halo (see section 5.6 of Roshan et al. 2021a). Another major issue is that the phonon-mediated forces at the heart of how SFDM reproduces the MOND phenomenology only enhance the forces on baryons, so it is difficult to understand why the RAR inferred

from rotation curves is also evident in strong lensing (Mistele, McGaugh & Hossenfelder 2022) and weak lensing data down to $g_N \approx 10^{-5} a_0$ (Brouwer et al. 2021) – the latter are particularly problematic for SFDM (Mistele, McGaugh & Hossenfelder 2023). These observational difficulties need to be considered alongside theoretical difficulties in making a stable covariant theory of DM superfluidity that reduces to MOND in galaxies (Hertzberg, Litterer & Shah 2021) and avoids significant orbital decay of stars due to Cherenkov-like radiation caused by their orbital velocity exceeding the local sound speed in the superfluid (Mistele 2022). If these difficulties are ultimately overcome, a positive aspect of SFDM is that the superfluid phase reduces to a normal phase near a star due to the steep potential gradient, avoiding any anomalous effects in the Solar System. However, this screening mechanism only works within a few hundred AU of a Sun-like star, beyond which MOND-like behaviour would be recovered (equation 86 of Berezhiani & Khoury 2015). Therefore, the WBT falsifies SFDM as presently understood. To pass the WBT, the force binding each WB should differ from the Newtonian expectation by $\lesssim 8$ per cent out to the separation limit of our WB sample. In the context of SFDM, this would require the normal phase bubble around every local Sun-like star to extend out to $\gtrsim 20$ kAU. This might be difficult to achieve by tuning the model parameters given constraints from galaxies and galaxy clusters (Hodson et al. 2017).

Thus, the extra forces needed to explain flat galaxy rotation curves might come from the mass of the DM rather than through it mediating a MOND-like interaction between the baryons. While the severe difficulties encountered by Λ CDM make this approach very unlikely (e.g. Banik & Zhao 2022, and references therein), the hypothetical DM particles might interact with each other – the standard assumption that they only interact gravitationally for all practical purposes is only made for simplicity. Self-interacting dark matter (SIDM; Spergel & Steinhardt 2000) offers a promising explanation for galaxies following a tight RAR (Ren et al. 2019). The required self-interaction cross-section is much larger than the upper limit imposed by the Newtonian dynamical mass profiles of galaxy clusters (Eckert et al. 2022b), but this could indicate that the cross-section depends strongly on the velocity. SIDM would limit the central DM density in galaxies, reducing dynamical friction on galaxy bars and thus possibly alleviating the 13σ fast bar tension faced by Λ CDM (Roshan et al. 2021b). SIDM would reduce to Λ CDM on large scales and thus still suffer the same issues with regards to the KBC void and Hubble tension (Haslbauer, Banik & Kroupa 2020) and the early formation of galaxies (Haslbauer et al. 2022) and galaxy clusters (Asencio, Banik & Kroupa 2021, 2023). In addition, SIDM does not provide an obvious explanation for the LG satellite planes because DM self-interactions do not have any obvious effect on the positions and velocities of satellite galaxies around their host – though the frequency of satellite planes in SIDM should be checked. Moreover, it is unclear how SIDM can explain the observed signs of tidal disturbance in Fornax Cluster dwarf galaxies and the lack of low surface brightness dwarfs towards the cluster centre (Asencio et al. 2022). Reducing the central DM density would make the situation less problematic, but the DM fraction within the baryonic extent of the Fornax dwarfs would need to be less than for isolated LG dwarfs with accurately measured internal velocity dispersions – these require a considerable amount of DM in a Newtonian context. Thus, SIDM models face some challenges but also appear to hold some promise, especially as they may explain some of the MOND phenomenology while being consistent with the WBT and Solar System ephemerides due to the lack of any change in the gravity sector. The same is true if we postulate a long-range non-

gravitational interaction between baryons and dark matter beyond their self-interactions (Famaey et al. 2020).

5.5 Future prospects

The significant challenge to MOND presented by the WBT is an important result which should be confirmed with additional data. There are good prospects for future improvement: the next *Gaia* data release (DR4) is anticipated at the end of 2025. DR4 will double the time baseline analysed in DR3, giving an $\approx 2.8\times$ improvement in proper motion precision and reducing the impact of CBs at AU separations (Appendix B). We estimate that using DR4 with a similar selection to our main analysis but extended to $m_G < 17.6$ and $d < 400$ pc would approximately double the search volume and almost triple the final sample size, bearing in mind that more precise proper motions will lead to more systems passing our cut on the \tilde{v} precision [equation [10]].

Systems rejected here due to a lack of either star’s RV can also be recovered from ground-based RV observations at a moderate precision of ≈ 1 km s $^{-1}$. This would increase the sample size by about 10 per cent, but the more useful gain could be that the contamination rate can be reduced if the RV is known for both stars.

There are also interesting prospects for directly constraining the triple population to reduce the degrees of freedom inherent in our modelling (Manchanda, Sutherland & Pittordis 2023). Their simulations have shown that we should be able to detect nearly all triples with a single main sequence star orbited by a CB whose inner period $\gtrsim 3$ years, which is long enough to give a substantial time-averaged perturbation to the photocentre velocity (Appendix B). Such hierarchical systems can be found either by *Gaia* astrometric accelerations in the 10-yr final data set (for separations $\lesssim 25$ AU); or by direct, speckle, or coronagraphic imaging (for separations $\gtrsim 20$ AU). The required imaging observations are much less time-consuming than high-precision RV measurements, so they may be feasible for a substantial fraction of the WB sample, especially for the high \tilde{v} tail. This would provide an important check on our estimated CB contamination fraction.

In principle, it is possible to refine the WBT via the addition of binary RV differences. This would allow the use of 3D velocities and 2D projected separations, which would reduce the scatter in the \tilde{v} distribution inherent from projected velocities. However, this would require high precision (order 0.05 km s $^{-1}$) RVs for all stars, which would be very costly in observing time and require corrections for gravitational redshift (Loeb 2022) and convective blueshift (Liebing et al. 2021). A fully 3D version of the WBT would also require much more precise astrometry to obtain kAU-level constraints on the relative heliocentric distances to the stars in each WB. This is not envisaged in the foreseeable future for a statistically large sample of WBs. More precise information is available for some very nearby systems (Kervella et al. 2016; Kervella, Thévenin & Lovis 2017), but an individual WB cannot test MOND using velocities alone even if $\tilde{v} > \sqrt{2}$ because a single snapshot would not prove that the WB is bound (though a chance flyby is unlikely; see section 8.1 of BZ18). A test is possible if the acceleration can be directly measured, which may become feasible with future observatories (Banik & Kroupa 2019a).

Our work has assumed that the potential of a star at kAU distances must be traced by another star. It is possible for a spacecraft to serve as the tracer instead (Banik & Kroupa 2019b). The predicted aspherical shell of PDM concentrated around the Solar MOND radius would also affect Solar System ephemerides (Hees et al. 2014, 2016; Brown & Mathur 2023).

6 SUMMARY AND CONCLUSIONS

MOND has enjoyed unparalleled predictive success with regards to galaxy dynamics (Famaey & McGaugh 2012; Banik & Zhao 2022, and references therein). Its central postulate is that the dynamics of a system deviates from Newtonian expectations when the gravity $g \lesssim a_0$. This could in principle occur at any distance from a system provided it has a sufficiently low mass [equation (2)]. Since the MOND radius of a star is $7\sqrt{M/M_\odot}$ kAU, MOND ought to have detectable effects on local WBs with kAU separations. Indeed, simple analytic estimates backed up by detailed calculations show a 20 per cent enhancement to the orbital velocity over the Newtonian expectation (BZ18), as also evident in N -body simulations of local star clusters for much the same reason (Kroupa et al. 2022).

We test this prediction using an observational sample of 8611 WBs from *Gaia* DR3 (Gaia Collaboration 2023) where the uncertainty on \tilde{v} [equation (3)] is very small and other quality cuts have been applied (Sections 2.1 and 2.4). Dividing our sample into ten equally sized subsamples shows no trend in the median \tilde{v} with respect to r_{sky}/r_M once we restrict our attention to $\tilde{v} \lesssim 2$, which should significantly reduce contaminating effects while preserving a genuine MOND signal (Fig. 11). Observational uncertainties would be expected to broaden the \tilde{v} distribution preferentially at large separations because the same uncertainty on the relative velocity implies a larger uncertainty in \tilde{v} . This would if anything lead to a broader \tilde{v} distribution at low accelerations. As a result, the flat behaviour of the median \tilde{v} with respect to our proxy for the internal WB acceleration strongly suggests that local WBs are Newtonian.

We then present the most detailed statistical hypothesis test of MOND to date using local WBs. Our model includes a rigorous calculation of the MOND gravitational field under an EFE of strength $g_e = 1.8 a_0$ (Section 3.1), an allowance for an undetected CB companion to one or both of the stars in each WB (Section 3.2), and LOS contamination (Section 3.3). The procedure was fixed in advance as much as possible to minimise biases that could arise when conducting the WBT (Banik, Pittordis & Sutherland 2021). The model parameters relate to those of:

- (i) The WBs, for which we need the α_{grav} parameter interpolating between different gravity laws [see equation (39) for its relation to the gravity law] and the distribution of semi-major axes [a_{break} and β ; see equation (16)] and orbital eccentricities [γ ; see equation (17)];
- (ii) The CBs, for which we assume the same γ parameter but also need the fraction f_{CB} of stars in our WB sample with an undetected CB companion (Section 3.2.4) and the upper limit to the CB semi-major axis, which is defined by k_{CB} [see equation (22)]; and
- (iii) The LOS contamination fraction f_{LOS} due to chance alignments (see Section 3.3).

These seven parameters are allowed to vary freely so as to best match the observed distribution of $(r_{\text{sky}}, \tilde{v})$ in 540 pixels (Table 1).

Our best-fitting Newtonian model significantly outperforms our best-fitting MOND model: the likelihood ratio of exp (175) implies a preference for Newtonian gravity at 19σ confidence according to our analyses with fixed α_{grav} . This is in line with the result of PS23, who conducted a less detailed version of the WBT using a somewhat different r_{sky} range of 5–20 kAU and considered WBs with $\tilde{v} < 7$. Their analysis also differs from ours in several other respects, including their use of a linear rather than cubic mass-luminosity relation (Fig. 1), their CB model, and their much more limited exploration of the parameter space. Those authors found that the χ^2 of their best Newtonian model is smaller than that of their best MOND model by 525, suggesting a 23σ preference for

Newtonian gravity. While our best-fitting Newtonian model is not a perfect representation of the WB data set, it is clear that the data do not show the predicted broadening to the \tilde{v} distribution at larger separations (Fig. 12).

In our main analysis, we allow the gravity law to interpolate between Newtonian and Milgromian. For this, we use the gravity law parameter α_{grav} , which is 0 in Newtonian gravity and 1 in MOND. Using 10^5 MCMC trials to explore the parameter space, we find that the inferred $\alpha_{\text{grav}} = -0.021^{+0.065}_{-0.045}$. This is very consistent with Newtonian dynamics but rules out MOND at 16σ confidence, which is in line with the results of PS23 and our results for fixed α_{grav} . Our result is robust to various changes in the modelling assumptions and the sample selection (Table 3), including when we use a narrower mass range to minimize possible trends in the mass distribution across the parameter range used for the WBT (Section 5.1.1). We also explain in some detail why, without a drastic change to our understanding of baryonic surface densities, the MOND interpolating function cannot be chosen to simultaneously pass the WBT and constraints from the rotation curves of galaxies, including our own (Section 5.3).

Our conclusion disagrees with two studies that were published while this work was under review (Chae 2023; Hernandez 2023). In Section 5.2, we focus on the study of Chae (2023) due to its much larger sample size and its claim to have detected the predicted MOND enhancement to Newtonian gravity at 10σ confidence. We find that the nominal sample of 26 615 WBs used in that study does indeed show a clear signal that closely resembles the MOND expectation and appears to rule out Newtonian dynamics (Fig. 19). However, we then identify a major deficiency with the handling of astrometric uncertainties. Despite the importance of the relative velocity between the stars in each WB, the uncertainty on this quantity is not estimated in Chae (2023). Instead, the focus is on ensuring that the heliocentric velocity of each star is very precise. This is insufficient to ensure a precisely known relative velocity and thus a reliable \tilde{v} , which however is essential to conducting the WBT. We therefore use equation 5 of El-Badry, Rix & Heintz (2021) to estimate the uncertainty in v_{sky} and thus in \tilde{v} . We then impose an additional condition on the \tilde{v} uncertainty matching that used in our own analysis [equation (10)]. This completely removes the apparent MOND signal despite only reducing the sample size by about 1/5 (see Fig. 21). We argue that a similar issue may well have affected the analysis of Hernandez (2023) based on some statements in that paper, though we do not study its WB sample in detail. We also find that the WB sample used by Chae (2023) artificially imposes that $\tilde{v} < \sqrt{5 M_\odot/M}$, making it not very well suited to the WBT (Appendix E).

We conclude that the gravity law inferred from our analysis of local WBs is consistent with Newtonian expectations but rules out MOND as modified gravity at $\gg 5\sigma$ confidence in both our nominal analysis and a considerable range of variations to it. This conclusion is of course reliant on our modelling approach, which is not a perfect match to the data (Fig. D1).¹⁹ Even so, neither our main analysis nor any of the variations considered can significantly improve the fit by using a gravity law different to Newtonian. This is also evident from the model-independent Fig. 11, which considers the median \tilde{v} in subsamples with different r_{sky}/r_M as a proxy for the WB internal

¹⁹This is similar to the rotation curve predictions in MOND, which often do not match the data within formal uncertainties but visually provide a good fit, which is typically considered sufficient given the inevitable modelling deficiencies (Kroupa et al. 2018; Cameron, Angus & Burgess 2020).

acceleration. It has been argued that modified inertia interpretations of MOND could reduce the predicted enhancement to the Newtonian acceleration and thus remain consistent with the WBT (Milgrom 1994, 2011, 2022). This would lead to the inner and outer parts of galaxy rotation curves following the same RAR, but a difference is detected at 6.9σ confidence (Chae 2022). Although the predictions of modified inertia for slightly non-circular orbits are not known, the magnitude of the above-mentioned difference is consistent with expectations of MOND as modified gravity – but this is ruled out by the WBT. Further modifications to MOND could preserve its successes on galaxy scales, especially with regards to issues like the LG satellite planes, which are difficult to understand any other way (Pawlowski 2021a, b; Banik et al. 2022). This entails at least one new fundamental scale beyond a_0 , marking the end of MOND as a purely acceleration-dependent modification to standard physics. Limiting the MOND phantom density to $|\rho_{\text{pdm}}| \lesssim 20 M_{\odot}/\text{pc}^3$ would yield consistency with the WBT and also alleviate tensions related to the null detection of MOND effects in Cassini radio tracking data from Saturn (Hees et al. 2014, 2016; Brown & Mathur 2023). It might also improve the agreement with observations of globular clusters and the vertical gravity from the Galactic disc. Limiting $|\rho_{\text{pdm}}|$ would have little effect on larger scales, similarly to the Vainshtein mechanism used to screen modified gravity effects in the Solar System (Babichev, Deffayet & Esposito-Farèse 2011). This would preserve MOND’s successes in galaxies but leave open the issue of galaxy clusters and large-scale structure, for which the hybrid ν HDM paradigm may be a promising approach (section 9.2 of Banik & Zhao 2022, and references therein). While our results falsify MOND as currently understood, given the many problems for Λ CDM discussed in that work, our results cannot be used to argue that it is the correct model either – both models are clearly incomplete. Hybrid models like SFDM struggle to explain galactic-scale observations like lensing (Mistele, McGaugh & Hossenfelder 2022, 2023) and the WBT, while non-MOND modifications to gravity like EG and MOG usually fail in galaxies (Section 5.4). SIDM may be a promising approach but it is very similar to Λ CDM on large scales, thus encountering the same difficulties with the KBC void (Keenan, Barger & Cowie 2013; Haslbauer, Banik & Kroupa 2020; Wong et al. 2022) and Hubble tension (Di Valentino 2021). This significant anomaly for standard cosmology appears to persist in the JWST era (Yuan et al. 2022) and must be solved consistently with the ages of the oldest stars (Cimatti & Moresco 2023, and references therein). We hope that our results from local WBs motivate the development of a more complete theory, which is likely to borrow some elements from both Λ CDM and MOND given their successes in different domains.

ACKNOWLEDGEMENTS

IB is supported by Science and Technology Facilities Council grant ST/V000861/1, which also partially supports HZ. IB acknowledges support from a “Pathways to Research” fellowship from the University of Bonn, during which the primary statistical analysis was largely coded. BF and RI acknowledge funding from the Agence Nationale de la Recherche (ANR projects ANR-18-CE31-0006 and ANR-19-CE31-0017) and from the European Research Council (ERC) under the European Union’s Horizon 2020 Framework programme (grant agreement number 834148).

The authors thank the referee for highlighting some important deficiencies with an earlier version of this work regarding the modelling of undetected companions. They also thank Pavel Kroupa for comments on the article and for lending computational resources from the Stellar Populations and Dynamics Research Group at the

University of Bonn, where all the statistical analyses were conducted. IB is very grateful to Elena Asencio for helping to check every line of the statistical analysis code and for lending algorithms used for making several figures in this paper, especially the triangle plots made using PYGTC (Bocquet & Carter 2016). He thanks Haixia Ma for providing Fig. 22. The authors are grateful for a masters project by Nathan Findlay which gave a much better understanding of how to analyse the data set and what quality cuts would be appropriate. IB acknowledges significant computational resources provided by Kyu-Hyun Chae to simulate the impact of undetected companions. Most of the figures were prepared using MATLAB.

DATA AVAILABILITY

The WB data set used in this contribution will be made available upon reasonable request, showing *Gaia* catalogue information for the sample of 19786 WBs described in the text and an extra flag to indicate if each WB is one of the 8611 systems used in our main analysis. The WB data set used in the nominal analysis of Chae (2023) was provided by Kyu-Hyun Chae. *Gaia* DR3 is publicly available.

REFERENCES

- Acedo L., 2020, *Universe*, 6, 209
 Ambartsumian V. A., 1937, *Astron. Zh.*, 14, 207
 Andrews J. J., Chanamé J., Agüeros M. A., 2017, *MNRAS*, 472, 675
 Angus G. W., 2009, *MNRAS*, 394, 527
 Angus G. W., Famaey B., Diaferio A., 2010, *MNRAS*, 402, 395
 Arenou F. et al., 2018, *A&A*, 616, A17
 Asencio E., Banik I., Kroupa P., 2021, *MNRAS*, 500, 5249
 Asencio E., Banik I., Mieske S., Venhola A., Kroupa P., Zhao H., 2022, *MNRAS*, 515, 2981
 Asencio E., Banik I., Kroupa P., 2023, *ApJ*, 954, 162
 Babichev E., Deffayet C., Ziour R., 2010, *Phys. Rev. D*, 82, 104008
 Babichev E., Deffayet C., Esposito-Farèse G., 2011, *Phys. Rev. D*, 84, 061502
 Bahcall J. N., Hut P., Tremaine S., 1985, *ApJ*, 290, 15
 Banik I., 2019, *MNRAS*, 487, 5291
 Banik I., Kroupa P., 2019a, *MNRAS*, 487, 1653
 Banik I., Kroupa P., 2019b, *MNRAS*, 487, 2665
 Banik I., Kroupa P., 2020a, *MNRAS*, 495, 3974
 Banik I., Kroupa P., 2020b, *MNRAS*, 497, L62
 Banik I., Zhao H., 2018a, *SciFed J. Astrophys.*, 1, 1000008
 Banik I., Zhao H., 2018b, *MNRAS*, 473, 419
 Banik I., Zhao H., 2018c, *MNRAS*, 480, 2660
 Banik I., Zhao H., 2022, *Symmetry*, 14, 1331
 Banik I., Milgrom M., Zhao H., 2018, preprint (arXiv:1808.10545)
 Banik I., Thies I., Candlish G., Famaey B., Ibata R., Kroupa P., 2020, *ApJ*, 905, 135
 Banik I., Pittordis C., Sutherland W., 2021, preprint (arXiv:2109.03827)
 Banik I., Thies I., Truelove R., Candlish G., Famaey B., Pawlowski M. S., Ibata R., Kroupa P., 2022, *MNRAS*, 513, 129
 Beech M., 2009, *MNRAS*, 399, L21
 Beech M., 2011, *Ap&SS*, 333, 419
 Begeman K. G., Broeils A. H., Sanders R. H., 1991, *MNRAS*, 249, 523
 Bekenstein J., Milgrom M., 1984, *ApJ*, 286, 7
 Belokurov V. et al., 2020, *MNRAS*, 496, 1922
 Bereziani L., Khoury J., 2015, *Phys. Rev. D*, 92, 103510
 Bereziani L., Elder B., Khoury J., 2019, *J. Cosmol. Astropart. Phys.*, 2019, 074
 Bilek M., Thies I., Kroupa P., Famaey B., 2018, *A&A*, 614, A59
 Bocquet S., Carter F. W., 2016, *J. Open Source Softw.*, 1, 46
 Bosma A., 1978, PhD thesis, Groningen University
 Bovy J., Bird J. C., García Pérez A. E., Majewski S. R., Nidever D. L., Zasowski G., 2015, *ApJ*, 800, 83
 Brouwer M. M. et al., 2021, *A&A*, 650, A113

- Brown K., Mathur H., 2023, *AJ*, 166, 168
- Cameron E., Angus G. W., Burgess J. M., 2020, *Nat. Astron.*, 4, 132
- Candlish G. N., Smith R., Jaffé Y., Cortesi A., 2018, *MNRAS*, 480, 5362
- Chabrier G., Baraffe I., 2000, *ARA&A*, 38, 337
- Chae K.-H., 2022, *ApJ*, 941, 55
- Chae K.-H., 2023, *ApJ*, 952, 128
- Chae K.-H., 2024, *ApJ*, in press (arXiv:2309.10404)
- Chae K.-H., Gong I.-T., 2015, *MNRAS*, 451, 1719
- Chae K.-H., Milgrom M., 2022, *ApJ*, 928, 24
- Chae K.-H., Bernardi M., Sheth R. K., Gong I.-T., 2019, *ApJ*, 877, 18
- Chae K.-H., Bernardi M., Domínguez Sánchez H., Sheth R. K., 2020a, *ApJ*, 903, L31
- Chae K.-H., Lelli F., Desmond H., McGaugh S. S., Li P., Schombert J. M., 2020b, *ApJ*, 904, 51
- Chae K.-H., Desmond H., Lelli F., McGaugh S. S., Schombert J. M., 2021, *ApJ*, 921, 104
- Chan M. H., Lee C. M., 2023, *MNRAS*, 518, 6238
- Chrobáková Ž., López-Corredoira M., Sylos Labini F., Wang H. F., Nagy R., 2020, *A&A*, 642, A95
- Cimatti A., Moresco M., 2023, *ApJ*, 953, 149
- Ciotti L., 2022, *ApJ*, 936, 180
- Clarke C. J., 2020, *MNRAS*, 491, L72
- Desmond H., 2023, *MNRAS*, 526, 3342
- de Almeida Á. O. F., Piattella O. F., Rodrigues D. C., 2016, *MNRAS*, 462, 2706
- de Salas P. F., Widmark A., 2021, *Rep. Prog. Phys.*, 84, 104901
- Di Valentino E., 2021, *MNRAS*, 502, 2065
- Dinnbier F., Kroupa P., Anderson R. I., 2022, *A&A*, 660, A61
- Eckert D., Ettori S., Pointecouteau E., van der Burg R. F. J., Loubser S. I., 2022a, *A&A*, 662, A123
- Eckert D., Ettori S., Robertson A., Massey R., Pointecouteau E., Harvey D., McCarthy I. G., 2022b, *A&A*, 666, A41
- Efstathiou G., Sutherland W. J., Maddox S. J., 1990, *Nature*, 348, 705
- Eilers A.-C., Hogg D. W., Rix H.-W., Ness M. K., 2019, *ApJ*, 871, 120
- Einstein A., 1915, *Sitzungsber. Kgl Preuss. Akad. Wiss. (Berlin)*, 6, p. 844
- Eker Z. et al., 2015, *AJ*, 149, 131
- El-Badry K., 2019, *MNRAS*, 482, 5018
- El-Badry K., Rix H.-W., 2018, *MNRAS*, 480, 4884
- El-Badry K., Rix H.-W., Tian H., Duchêne G., Moe M., 2019, *MNRAS*, 489, 5822
- El-Badry K., Rix H.-W., Heintz T. M., 2021, *MNRAS*, 506, 2269
- Faber S. M., Gallagher J. S., 1979, *ARA&A*, 17, 135
- Famaey B., Binney J., 2005, *MNRAS*, 363, 603
- Famaey B., McGaugh S. S., 2012, *Living Rev. Relativ.*, 15, 10
- Famaey B., Bruneton J.-P., Zhao H., 2007, *MNRAS*, 377, L79
- Famaey B., Khoury J., Penco R., Sharma A., 2020, *J. Cosmol. Astropart. Phys.*, 2020, 025
- Feng F., Jones H. R. A., 2018, *MNRAS*, 473, 3185
- Fletcher R., Powell M. J. D., 1963, *Comput. J.*, 6, 163
- Francis C., Anderson E., 2014, *Celest. Mech. Dyn. Astron.*, 118, 399
- Gaia Collaboration, 2016, *A&A*, 595, A1
- Gaia Collaboration, 2018, *A&A*, 616, A1
- Gaia Collaboration, 2021a, *A&A*, 649, A1
- Gaia Collaboration, 2021b, *A&A*, 649, A9
- Gaia Collaboration, 2023, *A&A*, 674, A1
- Gelman A., Gilks W. R., Roberts G. O., 1997, *Ann. Appl. Probab.*, 7, 110
- Gentile G., Famaey B., de Blok W. J. G., 2011, *A&A*, 527, A76
- Gordon D., de Witt A., Jacobs C. S., 2023, *AJ*, 165, 49
- Gravity Collaboration, 2019, *A&A*, 625, L10
- Hagen J. H. J., Helmi A., 2018, *A&A*, 615, A99
- Haghi H., Amiri V., Hasani Zonoozi A., Banik I., Kroupa P., Haslbauer M., 2019, *ApJ*, 884, L25
- Hajdukovic D. S., 2020, *MNRAS*, 491, 4816
- Hartman Z. D., Lépine S., Medan I., 2022, *ApJ*, 934, 72
- Haslbauer M., Banik I., Kroupa P., 2020, *MNRAS*, 499, 2845
- Haslbauer M., Kroupa P., Zonoozi A. H., Haghi H., 2022, *ApJ*, 939, L31
- Hayashi C., 1961, *PASJ*, 13, 450
- Hayashi C., Hoshi R., 1961, *PASJ*, 13, 442
- Hees A., Folkner W. M., Jacobson R. A., Park R. S., 2014, *Phys. Rev. D*, 89, 102002
- Hees A., Famaey B., Angus G. W., Gentile G., 2016, *MNRAS*, 455, 449
- Hees A., Famaey B., Bertone G., 2017, *Phys. Rev. D*, 95, 064019
- Hejlesen P. M., 1980, *A&AS*, 39, 347
- Hernandez X., 2023, *MNRAS*, 525, 1401
- Hernandez X., Jiménez M. A., Allen C., 2012, *Eur. Phys. J. C*, 72, 1884
- Hernandez X., Cortés R. A. M., Allen C., Scarpa R., 2019, *Int. J. Mod. Phys. D*, 28, 1950101
- Hernandez X., Cookson S., Cortés R. A. M., 2022, *MNRAS*, 509, 2304
- Hernandez X., Verteletskiy V., Nasser L., Aguayo-Ortiz A., 2024, *MNRAS*, in press (arXiv:2309.10995)
- Hertzberg M. P., Litterer J. A., Shah N., 2021, *J. Cosmol. Astropart. Phys.*, 2021, 015
- Hodson A. O., Zhao H., Khoury J., Famaey B., 2017, *A&A*, 607, A108
- Hoof S., Geringer-Sameth A., Trotta R., 2020, *J. Cosmol. Astropart. Phys.*, 2020, 012
- Hwang H.-C., Ting Y.-S., Zakamska N. L., 2022, *MNRAS*, 512, 3383
- Ibata R., Sollima A., Nipoti C., Bellazzini M., Chapman S. C., Dalessandro E., 2011a, *ApJ*, 738, 186
- Ibata R., Sollima A., Nipoti C., Bellazzini M., Chapman S. C., Dalessandro E., 2011b, *ApJ*, 743, 43
- Jao W.-C., Henry T. J., Gies D. R., Hambly N. C., 2018, *ApJ*, 861, L11
- Jeans J. H., 1919, *MNRAS*, 79, 408
- Jiang Y.-F., Tremaine S., 2010, *MNRAS*, 401, 977
- Jones-Smith K., Abraham R., Kell L., Mathur H., 2018, *New J. Phys.*, 20, 063042
- Kashfi T., Roshan M., Famaey B., 2023, *MNRAS*, 519, L21
- Keenan R. C., Barger A. J., Cowie L. L., 2013, *ApJ*, 775, 62
- Kervella P., Mignard F., Mérand A., Thévenin F., 2016, *A&A*, 594, A107
- Kervella P., Thévenin F., Lovis C., 2017, *A&A*, 598, L7
- Knox R. A., Hawkins M. R. S., Hambly N. C., 1999, *MNRAS*, 306, 736
- Kornetref C., Kaczmarek T., Pflanzner S., 2012, *A&A*, 543, A126
- Kroupa P., 2008, *Initial Conditions for Star Clusters. Lecture Notes in Physics Series*, Springer, Dordrecht, The Netherlands, p. 181
- Kroupa P., 2012, *Publ. Astron. Soc. Aust.*, 29, 395
- Kroupa P., 2015, *Can. J. Phys.*, 93, 169
- Kroupa P., Tout C. A., Gilmore G., 1993, *MNRAS*, 262, 545
- Kroupa P. et al., 2010, *A&A*, 523, A32
- Kroupa P. et al., 2018, *Nat. Astron.*, 2, 925
- Kroupa P. et al., 2022, *MNRAS*, 517, 3613
- LUX Collaboration, 2017, *Phys. Rev. Lett.*, 118, 021303
- Lelli F., McGaugh S. S., Schombert J. M., 2016, *AJ*, 152, 157
- Lelli F., McGaugh S. S., Schombert J. M., 2017a, *MNRAS*, 468, L68
- Lelli F., McGaugh S. S., Schombert J. M., Pawlowski M. S., 2017b, *ApJ*, 836, 152
- Lépine S., Bongiorno B., 2007, *AJ*, 133, 889
- Li P., Lelli F., McGaugh S., Schombert J., 2018, *A&A*, 615, A3
- Li P. et al., 2023, *A&A*, 677, A24
- Liebing F., Jeffers S. V., Reiners A., Zechmeister M., 2021, *A&A*, 654, A168
- Lindgren L. et al., 2021, *A&A*, 649, A4
- Loeb A., 2022, *Res. Notes AAS*, 6, 55
- Lux-Zeplin Collaboration, 2023, *Phys. Rev. Lett.*, 131, 041002
- Maeder A., Gueorguiev V. G., 2020, *MNRAS*, 492, 2698
- Manchanda D., Sutherland W., Pittordis C., 2023, *Open J. Astrophys.*, 6, E2
- Mann A. W., Feiden G. A., Gaidos E., Boyajian T., von Braun K., 2015, *ApJ*, 804, 64
- Mansfield S., Kroupa P., 2021, *A&A*, 650, A184
- McConnachie A. W., 2012, *AJ*, 144, 4
- McCulloch M. E., Lucio J. H., 2019, *Ap&SS*, 364, 121
- McGaugh S. S., 2008, *ApJ*, 683, 137
- McGaugh S. S., 2012, *AJ*, 143, 40
- McGaugh S. S., 2018, *Res. Notes AAS*, 2, 156
- McGaugh S., 2020, *Galaxies*, 8, 35
- McGaugh S. S., de Blok W. J. G., 1998a, *ApJ*, 499, 66
- McGaugh S. S., de Blok W. J. G., 1998b, *ApJ*, 499, 41
- McGaugh S. S., Schombert J. M., Bothun G. D., 1995, *AJ*, 109, 2019
- McGaugh S., Lelli F., Schombert J., 2016, *Phys. Rev. Lett.*, 117, 201101

- McGaugh S. S., Lelli F., Schombert J. M., Li P., Visgaitis T., Parker K. S., Pawlowski M. S., 2021, *AJ*, 162, 202
- McMillan P. J., 2017, *MNRAS*, 465, 76
- Migaszewski C., 2023, *MNRAS*, 525, 805
- Milgrom M., 1983, *ApJ*, 270, 365
- Milgrom M., 1986a, *ApJ*, 302, 617
- Milgrom M., 1986b, *ApJ*, 306, 9
- Milgrom M., 1994, *Ann. Phys.*, 229, 384
- Milgrom M., 2010, *MNRAS*, 403, 886
- Milgrom M., 2011, *Acta Phys. Pol. B*, 42, 2175
- Milgrom M., 2022, *Phys. Rev. D*, 106, 064060
- Milgrom M., 2023a, *Phys. Rev. D*, 108, 063009
- Milgrom M., 2023b, *Phys. Rev. D*, 108, 084005
- Mistele T., 2022, *J. Cosmol. Astropart. Phys.*, 2022, 008
- Mistele T., McGaugh S., Hossenfelder S., 2022, *A&A*, 664, A40
- Mistele T., McGaugh S., Hossenfelder S., 2023, *J. Cosmol. Astropart. Phys.*, 2023, 004
- Moffat J. W., 2006, *J. Cosmol. Astropart. Phys.*, 2006, 004
- Mróz P. et al., 2019, *ApJ*, 870, L10
- Nagesh S. T. et al., 2023, *MNRAS*, 519, 5128
- Negrelli C., Benito M., Landau S., Iocco F., Kraisselburd L., 2018, *Phys. Rev. D*, 98, 104061
- Nelson T., Ting Y.-S., Hawkins K., Ji A., Kamdar H., El-Badry K., 2021, *ApJ*, 921, 118
- Offner S. S. R., Moe M., Kratter K. M., Sadavoy S. I., Jensen E. L. N., Tobin J. J., 2023, in Inutsuka S., Aikawa Y., Muto T., Tomida K., Tamura M. eds, *ASP Conf. Ser. Vol. 534, Protostars and Planets VII*, Astron. Soc. Pac., San Francisco, p. 275
- Oh S., Price-Whelan A. M., Hogg D. W., Morton T. D., Spergel D. N., 2017, *AJ*, 153, 257
- Õpik E., 1924, *Publ. Tartu Astros. Obs.*, 25, 1
- Oria P. A. et al., 2021, *ApJ*, 923, 68
- Ostriker J. P., Steinhardt P. J., 1995, *Nature*, 377, 600
- Pace A. B., Erkal D., Li T. S., 2022, *ApJ*, 940, 136
- Paučo R., Klačka J., 2016, *A&A*, 589, A63
- Pawlowski M. S., 2021a, *Nat. Astron.*, 5, 1185
- Pawlowski M. S., 2021b, *Galaxies*, 9, 66
- Pecaut M. J., Mamajek E. E., 2013, *ApJS*, 208, 9
- Peebles P. J. E., 2017, *Nat. Astron.*, 1, 0057
- Penner A. R., 2020, *Ap&SS*, 365, 65
- Penoyre Z., Belokurov V., Evans N. W., 2022, *MNRAS*, 513, 2437
- Petersen J., Lelli F., 2020, *A&A*, 636, A56
- Pflamm-Altenburg J., Kroupa P., Thies I., Jerabkova T., Beccari G., Prusti T., Boffin H. M. J., 2023, *A&A*, 671, A88
- Pichon B., 2007, in Bouvier J., Chababae A., Charbonnel C. eds, *SF2A-2007: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*. Grenoble, France, Soci'et'e Francaise d'Astronomie et d'Astrophysique, p. 549
- Pittordis C., Sutherland W., 2018, *MNRAS*, 480, 1778
- Pittordis C., Sutherland W., 2019, *MNRAS*, 488, 4740
- Pittordis C., Sutherland W., 2023, *Open J. Astrophys.*, 6, 4
- Planck Collaboration VI, 2020, *A&A*, 641, A6
- Read J. I., 2014, *J. Phys. G Nucl. Phys.*, 41, 063101
- Reid M. J., Brunthaler A., 2004, *ApJ*, 616, 872
- Reid I. N., Gizis J. E., Hawley S. L., 2002, *AJ*, 124, 2721
- Ren T., Kwa A., Kaplinghat M., Yu H.-B., 2019, *Phys. Rev. X*, 9, 031020
- Renaud F., Famaey B., Kroupa P., 2016, *MNRAS*, 463, 3637
- Riddle R. L. et al., 2015, *ApJ*, 799, 4
- Riello M. et al., 2021, *A&A*, 649, A3
- Roshan M., Banik I., Ghafourian N., Thies I., Famaey B., Asencio E., Kroupa P., 2021a, *MNRAS*, 503, 2833
- Roshan M., Ghafourian N., Kashfi T., Banik I., Haslbauer M., Cuomo V., Famaey B., Kroupa P., 2021b, *MNRAS*, 508, 926
- Rowland D. R., 2015, *Int. J. Mod. Phys. D*, 24, 1550065
- Rubin V. C., Ford W. K. J., Thonnard N., 1978, *ApJ*, 225, L107
- Salomon J.-B., Bienaymé O., Reylic C., Robin A. C., Famaey B., 2020, *A&A*, 643, A75
- Sanders R. H., 2003, *MNRAS*, 342, 901
- Sanders R. H., 2012a, *MNRAS*, 419, L6
- Sanders R. H., 2012b, *MNRAS*, 422, L21
- Schlegel D. J., Finkbeiner D. P., Davis M., 1998, *ApJ*, 500, 525
- Schombert J., McGaugh S., Lelli F., 2022, *AJ*, 163, 154
- Schönrich R., 2012, *MNRAS*, 427, 274
- Schönrich R., Binney J., Dehnen W., 2010, *MNRAS*, 403, 1829
- Sellwood J. A., Shen J., Li Z., 2019, *MNRAS*, 486, 4710
- Shaya E. J., Olling R. P., 2011, *ApJS*, 192, 2
- Skordis C., Złośnik T., 2019, *Phys. Rev. D*, 100, 104013
- Skordis C., Złośnik T., 2021, *Phys. Rev. Lett.*, 127, 161302
- Skordis C., Złośnik T., 2022, *Phys. Rev. D*, 106, 104041
- Spergel D. N., Steinhardt P. J., 2000, *Phys. Rev. Lett.*, 84, 3760
- Tamosiunas A., Bacon D., Koyama K., Nichol R. C., 2019, *J. Cosmol. Astropart. Phys.*, 2019, 053
- Thomas G. F., Famaey B., Ibata R., Renaud F., Martin N. F., Kroupa P., 2018, *A&A*, 609, A44
- Thomas G. F., Famaey B., Monari G., Laporte C. F. P., Ibata R., de Laverny P., Hill V., Boily C., 2023, *A&A*, 678, A180
- Tian Y., Ko C.-M., 2019, *MNRAS*, 488, L41
- Tokovinin A., 2017, *MNRAS*, 468, 3461
- Tokovinin A., 2023, *AJ*, 165, 165
- Tokovinin A., Moe M., 2020, *MNRAS*, 491, 5158
- Tokovinin A., Smekhov M. G., 2002, *A&A*, 382, 118
- Tokovinin A., Hartung M., Hayward T. L., 2010, *AJ*, 140, 510
- Vainshtein A. I., 1972, *Phys. Lett. B*, 39, 393
- Verlinde E., 2017, *SciPost Phys.*, 2, 016
- Wang J., Hammer F., Yang Y., 2022, *MNRAS*, 510, 2242
- Wong J. H. W., Shanks T., Metcalfe N., Whitbourn J. R., 2022, *MNRAS*, 511, 5742
- Yuan W., Riess A. G., Casertano S., Macri L. M., 2022, *ApJ*, 940, L17
- Zhao H., 2007, *ApJ*, 671, L1
- Zhao H., Li B., Bienaymé O., 2010, *Phys. Rev. D*, 82, 103001
- Zhao H., Famaey B., Lüghausen F., Kroupa P., 2013, *A&A*, 557, L3
- Zhou Y., Li X., Huang Y., Zhang H., 2023, *ApJ*, 946, 73
- Zhu Y. et al., 2023, *MNRAS*, 519, 4479
- Zonoozi A. H., Lieberz P., Banik I., Haghi H., Kroupa P., 2021, *MNRAS*, 506, 5468

APPENDIX A: EXACTLY EQUAL MASS BINARIES

Although we might expect the mass ratio distribution of binaries to be smooth, observations indicate a population of exactly equal mass WBs (El-Badry et al. 2019). We therefore include an allowance for a δ -function in the \tilde{q} distribution at $1/2$, where \tilde{q} is the fraction of the CB mass in the less massive star. To find the likelihood P_{eqm} of a binary having exactly equal mass stars, we plot the cumulative distribution of $0.5 - \tilde{q}$ for the WBs in our sample, which we restrict to $\tilde{v} < 1$ to try and mitigate CB contamination. The idea is to extrapolate the distribution down to zero and look for a positive intercept. Since a δ -function would be smeared somewhat by observational uncertainties, we need to start our fit at some slightly positive value of $0.5 - \tilde{q}$. We also need to avoid extending our fit to very high $0.5 - \tilde{q}$ because our polynomial fitting function might become inaccurate.

The results in the top panel of Fig. A1 show real data and a cubic fit over the range $0.05 - 0.2$. The data are shown in black, while the cubic fit is shown in red – a cubic is the lowest degree polynomial which provides an accurate match to the data. The bottom panel of Fig. A1 shows the intercept as a function of the upper limit to $0.5 - \tilde{q}$, with each coloured line used to show results for a different lower limit. All analysis variants show a positive intercept, whose value converges at close to 0.04 for a wide range of lower and upper limits to $0.5 - \tilde{q}$. We therefore adopt $P_{\text{eqm}} = 0.04$ for the CB model in our main analysis. This assumption is not relevant to modelling uncontaminated WBs, whose actual mass ratio distribution is used

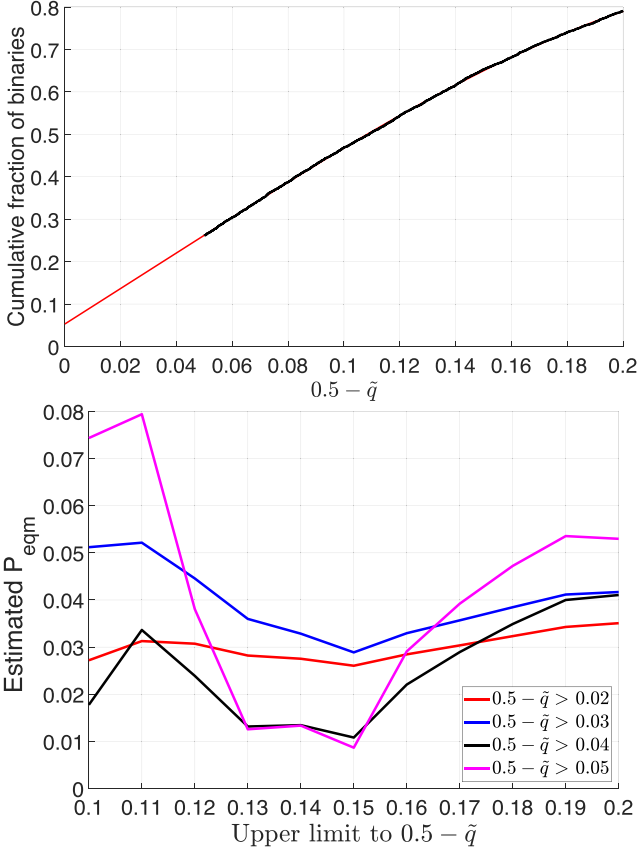


Figure A1. *Top:* The cumulative distribution of $0.5 - \tilde{q}$ for WBs with $\tilde{v} < 1$ (black line). We only show results over the range $0.05 - 0.2$ because a δ -function in the probability distribution at zero would be smeared somewhat by observational uncertainties. The red line shows a cubic fit over this range, but also extrapolates it down to zero. Notice that the intercept is slightly positive, indicating the presence of a population of WBs with an exactly equal mass (consistent with El-Badry et al. 2019). *Bottom:* The above-mentioned intercept is plotted as a function of the upper limit to the fitting range. Different lines show results with different lower limits, as indicated in the legend. Notice that for a wide range of fitting ranges, the results converge around 0.04, which we assume is the likelihood P_{eqm} of a binary having exactly equal mass components.

when needed (Section 3.2.4). An equal mass fraction of a few percent is in line with the double main sequence apparent in the colour-magnitude diagram of our WB sample (Fig. 2).

APPENDIX B: WHEN ORBITAL MOTION BECOMES NON-LINEAR

We assume uniform rectilinear motion for the WB and any undetected CB companions, implicitly assuming that the orbital periods are very long. This should be a very good approximation for the WB given the kAU separations of even the tightest WBs we consider. However, CB orbital periods can be much shorter. This can substantially reduce the mean motion of the CB over the *Gaia* DR3 observing baseline of $t_G = 34$ months. In this section, we estimate the CB orbital period below which we can no longer safely assume a very long orbital period P . This is used to estimate a minimum CB semi-major axis a_{int} when setting up its distribution in Section 3.2.3.

For simplicity, we assume that the CB is on a circular orbit and that the line connecting its components rotates by some angle θ over

the *Gaia* observing baseline. We consider the impact of increasing θ by reducing the CB separation while not altering the other CB orbital parameters. The induced recoil velocity on the star detected as part of the WB depends on the CB orbital velocity $v \propto 1/\sqrt{a_{\text{int}}}$. Bearing in mind Kepler’s Third Law that $P \propto a_{\text{int}}^{3/2}$, we get that $v \propto P^{-1/3}$. Since the CB rotation angle θ over a fixed duration of time varies as $\theta \propto 1/P$, we get that

$$v \propto \theta^{1/3}. \quad (\text{B1})$$

The positive exponent captures the fact that larger θ is associated with a tighter CB on a faster orbit.

However, the mean velocity \bar{v} will in general involve some additional shape factor $s(\theta)$ that accounts for the extent to which the orbital arc over the *Gaia* observing baseline deviates from a straight line. Basic trigonometry tells us that the linear distance between two points on a circle is smaller than the distance between them along the circumference by a factor of $\text{sinc}(\theta/2)$. Putting this extra factor into equation (B1) tells us that

$$\bar{v} \propto \theta^{-2/3} \left| \sin\left(\frac{\theta}{2}\right) \right|. \quad (\text{B2})$$

Treating this as a function of $u \equiv \theta/2$, we find that the maximum occurs when $\tan u = 3u/2$. The first non-trivial solution is $u = 0.97$ or $\theta = 111^\circ$. Rotation by this angle over t_G implies $P = 3.25 t_G$, so we expect our linear motion approximation to break down when $P < 110$ months or 9.2 yr.

The corresponding orbital separation depends on the mass of the CB. For a low-mass undetected companion, the CB total mass would only slightly exceed the mass of the contaminated star, which is just one of the two stars detected as a WB. The blue bars in Fig. 5 show that the stars in our WB sample easily reach down to $0.4 M_\odot$. At this mass, an orbital period of 9.2 years corresponds to $a_{\text{int}} = 3.2$ AU. While CBs with even smaller separations would still affect the inferred WB orbital velocity, the impact rapidly becomes smaller than implied by equation (B1).

We note that 3.2 AU is much larger than the minimum a_{int} of 0.1 AU considered in the orbit integrations of PS23 (see their section 3.2). CBs with a separation of 0.1 AU would complete many orbits over t_G , so the effect of such tight CBs may have been overestimated in their analysis.

APPENDIX C: COMPARING THE BEST NEWTONIAN AND MOND MODELS

In Fig. 13, we presented the difference in log-likelihood between our best-fitting Newtonian and Milgromian models. The panels in this figure showed the results for each r_{sky} but with all \tilde{v} pixels summed over, and vice versa. To give a better understanding of which pixels work better in which theory, we show the full 2D distribution of $\Delta \ln P$ between these models in Fig. C1. Since $\Delta \ln P$ is close to zero in several pixels, we clarify which model does better by adding an open white circle at the centre of a pixel if MOND fits it better. Out of the 540 pixels used in our analysis (Table 1), only 221 pixels (41 per cent) prefer MOND while the remaining 59 per cent prefer Newtonian gravity. However, we also need to consider that the pixels which prefer MOND only do so to a rather small extent, while sometimes pixels which prefer Newtonian gravity do so to a very large extent.

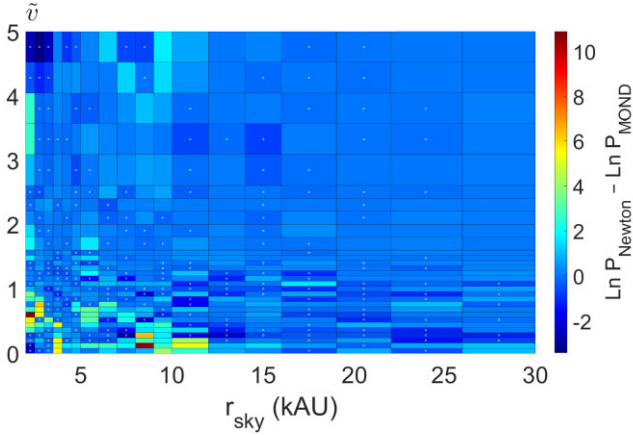


Figure C1. The relative performance of our best-fitting Newtonian and MOND models, shown here for every pixel. The binomial log-likelihood of each model is found by comparison with the observed number of WBs in each pixel (Fig. 6). Open white circles indicate pixels for which MOND outperforms Newtonian gravity. The combined difference in log-likelihood at each r_{sky} summed across all \tilde{v} (and vice versa) is shown in Fig. 13.

APPENDIX D: THE BEST FIT MODEL WITH REDUCED CB CONTAMINATION

The only alteration to our nominal analysis which gives an appreciable preference for $\alpha_{\text{grav}} > 0$ is the one in which we fix $f_{\text{CB}} = 0.3$ (Table 3). We explore this model in more detail to understand if it plausibly fits the WB data with a gravity law closer to Milgromian.

In Fig. D1, we compare the observed \tilde{v} distribution to the prediction of the best model with our nominal assumptions and with $f_{\text{CB}} = 0.3$, considering both the final result of the gradient ascent and the whole MCMC chain in each case. It is clear that fixing f_{CB} to such a low value causes a catastrophic disagreement with nearly all aspects of the observations. The lack of sufficient CB contamination causes the peak of the \tilde{v} distribution to be more pronounced than in the Gaia number counts at very low r_{sky} , though we see that both models struggle somewhat in the peak region at intermediate r_{sky} . The most important differences concern the tail of the distribution, which cannot be adequately fit with so few CBs. The analysis has to try and fit the extended tail with substantially more LOS contamination, quadrupling the inferred f_{LOS} . This may help somewhat at low r_{sky} , but it leads to a rapidly rising \tilde{v} distribution at high r_{sky} , in strong disagreement with the observations (bottom right panel).

These very serious issues cause the overall fit to be poorer than in our best-fitting nominal analysis by $\Delta \ln P = 413$, which implies that the latter is preferred at a confidence equivalent to $\sqrt{2\Delta \ln P} = 29\sigma$ for a 1D Gaussian. Clearly, assuming a substantially lower likelihood of CB contamination is not a viable proposition. In any case, this only shifts the inferred gravity law towards MOND by a small amount, with Newtonian gravity still strongly preferred despite facing just over 3σ tension (Fig. 18). It is therefore extremely difficult to reconcile our WB results with MOND as usually understood given other constraints.

APPENDIX E: THE ARTIFICIAL VELOCITY LIMIT IN THE CHAE (2023) ANALYSIS

In Section 5.2, we discussed several problems with the WB analysis of Chae (2023). One additional problem which may be noteworthy is that the sample is based on the WB catalogue of El-Badry, Rix & Heintz (2021). Each WB is assigned a likelihood of being genuine based on several factors, one of the most important being the magnitude of v_{sky} (see their equation 7).²⁰ Unlike our limit that $\tilde{v} < 5$, their limit to v_{sky} does not easily translate into a limit on \tilde{v} because there is no allowance for more massive WBs having a faster Newtonian v_c . Those authors require that

$$v_{\text{sky}} < 2.1 \sqrt{\frac{\text{kAU}}{r_{\text{sky}}}} \text{ km s}^{-1}, \quad (\text{E1})$$

which in terms of \tilde{v} corresponds to

$$\tilde{v} < \sqrt{\frac{5 M_{\odot}}{M}}. \quad (\text{E2})$$

We illustrate this cutoff in Fig. E1, where we plot \tilde{v} as a function of M for the 26615 WBs in the nominal sample of Chae (2023). Nearly all of these WBs lie below the blue line showing the limit given by equation (E2). This is not a completely strict limit because each WB is required to be consistent with this bound within the 2σ observational uncertainty on v_{sky} (see section 2 of El-Badry, Rix & Heintz 2021). Even so, it should be clear that a WB sample selected in this way is not ideal for the WBT.²¹

To check if our results in Section 5.2 remain reliable despite this complication, we need to impose an upper limit on the mass so that results remain reliable up to some maximal \tilde{v} . This is equivalent to drawing a rectangular cut towards the bottom left of Fig. E1. To retain most of the WB sample while having reliable results beyond the main peak of the \tilde{v} distribution, we impose a limit of $M < 1.25 M_{\odot}$. With this mass limit, regardless of the mass of a WB, there is no truncation to the \tilde{v} distribution for $\tilde{v} < 2$. We then use Fig. E2 to show the median \tilde{v} as a function of r_{sky}/r_M for the case where the sample is further limited to only those WBs where the estimated $\sigma(\tilde{v}) < 0.1 \max(1, \tilde{v}/2)$, as done in Fig. 21. Both figures show very similar results, indicating that the \tilde{v} limit imposed by equation (E2) has little impact on the flat trend of the median \tilde{v} with respect to our proxy for the internal WB acceleration. This is presumably because the main peak of the \tilde{v} distribution lies at $\tilde{v} < 1$, which is not much affected by this cut (Fig. E1). Even so, having a mass-dependent limit to \tilde{v} could in principle bias the WBT because a WB with a lower mass also has a lower acceleration at the same r_{sky} . Since the main idea of the WBT is that some typical measure of $\tilde{v} \propto \sqrt{g/g_N}$, one should be cautious about a sample selection that allows lower mass systems to reach a higher \tilde{v} than higher mass systems.

²⁰This is estimated using the difference in proper motions without allowing for the systemic RV, which should be a reasonable approximation given the small angular sizes of WBs.

²¹K. El-Badry, private communication.

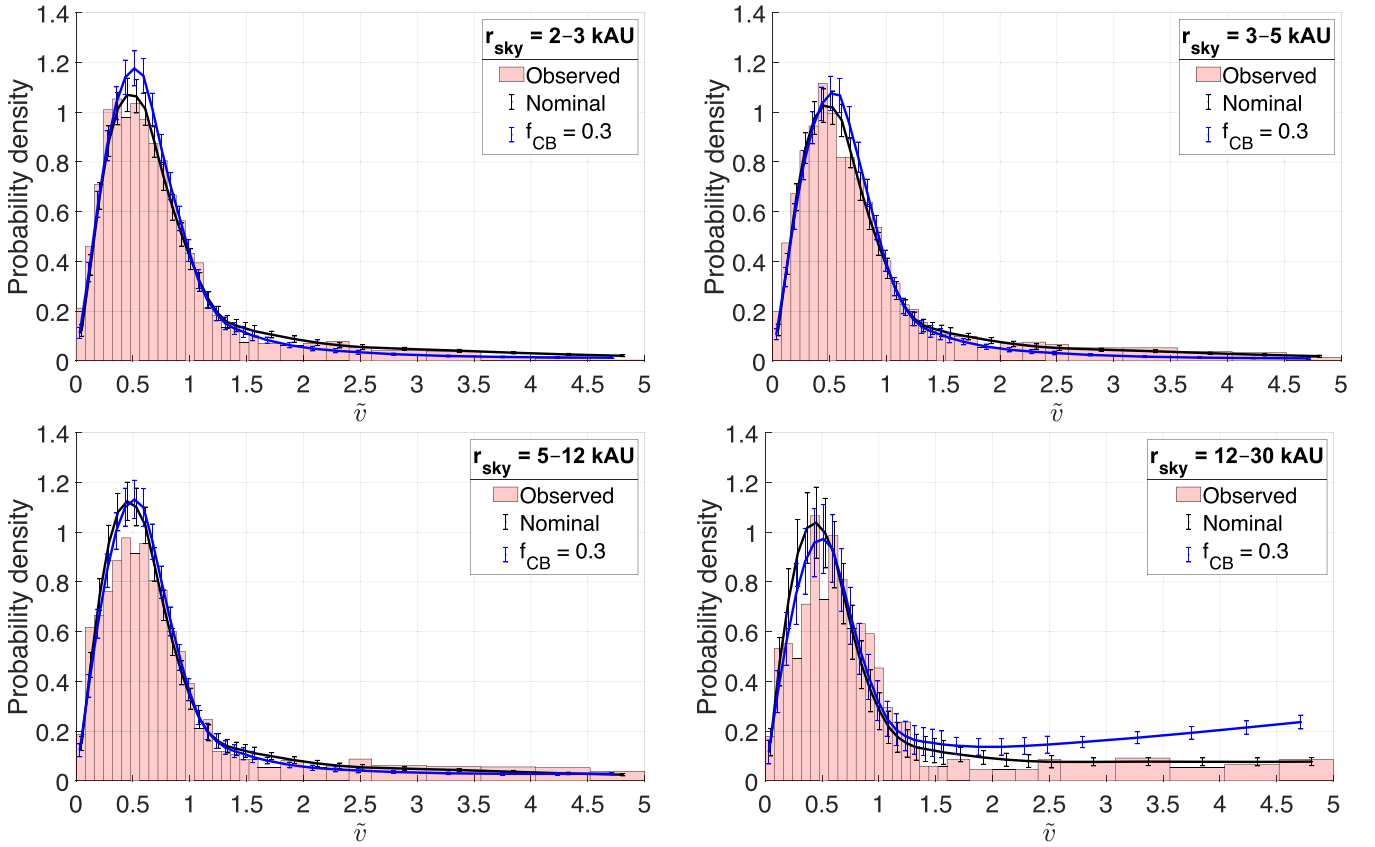


Figure D1. Comparison between the observed \tilde{v} distribution in four different r_{sky} ranges (solid red bars in different panels) with the prediction of our best model from our nominal analysis (black lines with error bars) and from our revised analysis with fixed $f_{\text{CB}} = 0.3$ (blue lines with error bars). This revised analysis is the only one to appreciably shift the inferred gravity law towards MOND (Fig. 18). The fit is much poorer in this case. The results shown here are normalized based on the observed WB distribution, helping to highlight that the $f_{\text{CB}} = 0.3$ model predicts too many WBs at high r_{sky} .

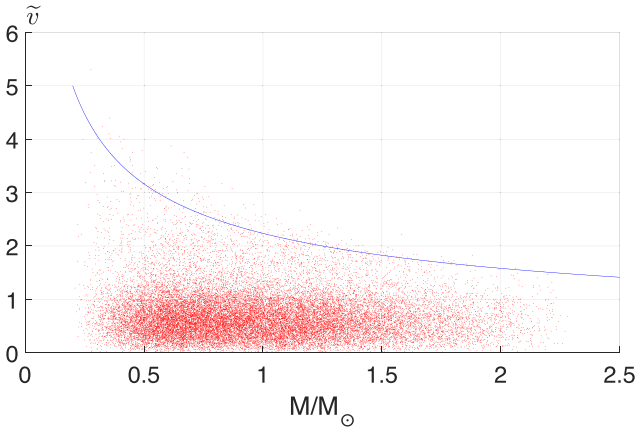


Figure E1. The red dots show the distribution of WB mass and \tilde{v} for the nominal sample of Chae (2023). Notice how the vast majority of WBs lie below the solid blue line [equation (E2)] due to selection effects (equation 7 of El-Badry, Rix & Heintz 2021).

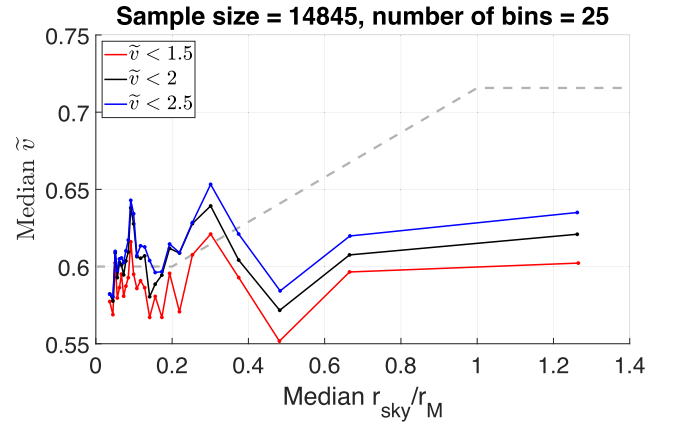


Figure E2. Similar to Fig. 21, but restricted further to only WBs with $M < 1.25 M_{\odot}$. This ensures that the artificial limit to the \tilde{v} distribution in the Chae (2023) sample (blue line in Fig. E1) does not influence the results for $\tilde{v} < 2$. The mass limit has little impact on the results, which continue to show a flat trend once we impose that $\sigma(\tilde{v}) < 0.1 \max(1, \tilde{v}/2)$, as done here.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.