Review

Galactic Anomalies and Particle Dark Matter

Malcolm Fairbairn

Department of Physics, King’s College London, Strand, London WC2R 2LS, UK; malcolm.fairbairn@kcl.ac.uk

Abstract: This is a brief review of aspects of galactic astrophysics and astronomy which have a possible bearing on particle dark matter. It is still quite normal for particle physicists to try to solve “well known anomalies” that are apparently seen in observations of galaxies (missing satellites, cusp vs. core, etc.) whereas a lot of these anomalies have actually been resolved many years ago. We will try to briefly review the field and discuss many of the areas in question.

Keywords: dark matter; galaxies; particle physics

1. Introduction

1.1. Evidence for Dark Matter

The journey to uncover the particle nature of dark matter (DM) is an ongoing process at the intersection of Astrophysics, Particle Physics, and Cosmology. State-of-the-art observations of the Cosmic Microwave Background from Planck [1] show that 26% of the energy density of the Universe consists of a form of matter that appears to interact gravitationally, but only at most very weakly, with other standard model particles. The substance either has no or very little pressure due to interactions with itself because it clusters in conditions where pressure prevents baryons from doing so (although it might have some self-interaction and some weak pressure). Furthermore, it emerges from the early Universe with a small enough velocity to become gravitationally bound [2]. Studies also universally suggest that galaxies and clusters of galaxies are dominated by some matter that does not emit light—see, e.g., references [3–8].

Observations of galaxy clustering [9], supernova redshift surveys [10], and the amplitude of the Cosmic Microwave Background (CMB) correlation function as a function of angular scale [1] and are all well explained in the $\Lambda$CDM framework [11] (a notable caveat to this is the current $H_0$ discrepancy [12]). The paradigm of cold dark matter in the context of the $\Lambda$CDM model describes dark matter as a non-relativistic, collisionless fluid which only couples gravitationally to the baryonic sector. $\Lambda$CDM provides an excellent theoretical description of the way matter is distributed on large scales and predicts hierarchical structure formation, namely that small structures form first and subsequently coalesce to form consecutively larger halos—this paradigm is backed up by many different observations [1,13,14].

1.2. Candidates for Dark Matter

A vast number of particle dark matter candidates have been suggested. The landscape of dark matter candidates extends in mass from $m \sim 10^{-22}$ eV for ultralight dark matter [15] to $m \sim 100 M_\odot$ in the case of primordial black holes [16]. Along the way between these two extremes, there is an array of well-motivated dark matter candidates, such as WIMPs/thermal relics with $m \sim 1$ MeV–100 TeV [17], axions (which if they are a dark matter particle typically have masses around 0.1 meV but may have a huge range of masses) [18], sterile neutrinos with $m \sim (1–100)$ keV [19], asymmetric dark matter with $m \sim 1$ GeV [20], and many more.

Since the time when particle dark matter candidates were proposed some decades ago, a large and extensive experimental programme has been underway to search for them. The
majority of this search was for a long time focused on WIMPs/thermal relics since it was thought that the weak-scale interactions responsible for setting their thermal abundance in the early Universe might be associated with new physics at the electroweak scale, which might be probed at the Large Hadron Collider. Searches for WIMPs are typically categorised as either make—creating dark matter from interactions at colliders, break—looking for dark matter particles that annihilate with each other in dense regions (such as at the centre of the galaxy, in dwarf galaxies and sub-halos, and in other regions where the density of dark matter is high, such as the early Universe), and shake—waiting for dark matter to interact with sensitive detectors lying underground. No definitive direct evidence for the existence of DM has been obtained so far from terrestrial [21–23], astrophysical [23–26], or collider [27,28] searches. There are some indications of excess gamma rays from the centre of the galaxy [29,30], but these are not universally accepted and a long running and detailed debate has been rolling on about whether or not this is due to millisecond pulsars or not [31–38].

1.3. Pinning down the Properties of Dark Matter Using Observations

It is important that people continue to search for dark matter using any available method but it is conceivable that it could be many years, decades, or even centuries before we detect it in a laboratory. It could be that dark matter particles only interact gravitationally. Certainly, a couple of decades ago it was very fashionable to go down these kinds of routes through string-inspired brane-world theories, where only gravity travelled between different gauge sectors separated on membranes in a compact space [39]. Furthermore, work carried out by the author and collaborators showed that there are a variety of models which are consistent with only very weak gravitational couplings to dark matter, particularly if it is non-minimally coupled and can therefore be produced copiously immediately after inflation [40].

This means that in order to find out more about the nature of this mysterious substance, we need to look into its distribution in space to see what it is doing [41]. This is a sure-fire way of studying dark matter, even if we never see it annihilate into, or interact with, standard model particles. This method is of course limited by our knowledge of the astrophysical systematics which dress those observations—we study the dark matter by its gravitational effects on the baryonic matter but we need to separate the effects on baryons due to baryons and those due to gravity. There are a number of ways to probe the nature of dark matter in this way, and by making these observations, we hope to answer some questions, including but not limited to:

- Is the dark matter cold or warm?
- Is the dark matter self-interacting? (Does it have some pressure?)
- Can we put a lower limit on the mass of the dark matter due it (not-)exhibiting particular fermionic or bosonic behaviour?

It is often implied that because we have been unable to detect particle dark matter, it is less likely to be a particle and more likely that the observations are due to some modification of gravity. The author does not really understand this train of logic, especially given the possible extremely weak nature of the interactions between dark matter and the visible sector. Nevertheless, we are also duty-bound to address the deeper question:

- Are the astrophysical observations consistent with dark matter being some kind of particle, or are they telling us something else?

The $\Lambda$CDM model is hugely successful at explaining the Universe we live in. As physicists, we know that we can never prove a theoretical interpretation of data, we can only rule it out. In that context, one interpretation of the scientific method is that attack is the highest form of flattery—anybody connected to trying to understand dark matter or even strengthen the case for $\Lambda$CDM needs to actively try to identify any cracks in that scenario and try to hurt it. The more such attacks it survives, the stronger it becomes as a paradigm. Over the years, several possible areas of tension between $\Lambda$CDM and
observations have been identified. It is not clear if these are real problems, and some of them have already gone away as we shall see. We will mention some of them below.

Interestingly, some of the problems which have gone away as far as many astrophysicists are concerned (e.g., the missing satellites problem [42–44]) or which are more complicated than people who do not look into galactic physics usually understand (e.g., the core vs. cusp problem [45–49]) are very often used by particle phenomenologists as motivation for some non-CDM properties which can be explained by their latest favourite particle physics Lagrangian. A phenomenological approach to dark matter, including particle and astrophysical data to constrain new models, is something the author is enthusiastic to support. However, one should be aware of the strengths and the weaknesses in the astrophysical anomalies one uses to constrain ones’ model.

2. Small-Scale Structures and Implications for Particle Physics

The hierarchical structure formation associated with cold dark matter means that smaller halos form first. Larger dark matter halos, such as galaxies, will therefore be composed of many smaller halos with progressively smaller masses, spanning very many orders of magnitude [50,51].

Alternatives to CDM, such as warm DM [52–54], self-interacting DM [55], and fuzzy DM [15,56], behave differently on small scales [1]. Such models possess a lack of smaller galaxies since dark matter halos would be unable to form below characteristic scales which have a different physical origin in each case (see Figure 1). In the case of warm dark matter, this corresponds to the free-streaming scale, the scale at which the dark matter has travelled at the moment it become non-relativistic [53]. There are effects on the halo mass function for ultralight or fuzzy dark matter due to their quantum nature changing the Jeans length in the dark sector that also suppresses structures on small scales in a somewhat similar way [57,58] (the two different effects could be disentangled with extremely good data in the future). The effect of self-interacting dark matter upon the number of smaller-mass sub-halos is less clear [59]. It is extremely important to identify whether smaller-mass dark matter halos exist, since this can distinguish between different models.

![Figure 1](image-url)

Figure 1. Diagram of how the power spectrum changes in the presence of warm dark matter. In cold dark matter, there are increasingly large numbers of smaller halos as you go down in mass (large $k$). As the dark matter becomes warmer, corresponding to a lighter particle if it is in thermal equilibrium with the plasma at some early time, it has enough velocity to wipe out structures at higher $k$, and subsequently the overdensities from regions smaller than those scales are wiped out. Smaller halos are therefore more scarce (this is Figure 1 taken from the paper [60] in accordance with the usage permissions of Oxford University Press who retain the copyright).
We can only measure dark matter through its gravitational effect upon stars and gas, so probes of its properties exist only from halos with a mass equal to or above the smallest galaxies containing stars. The dark-matter-dominated low-mass end of the halo mass function is challenging to probe due to a lack of stars—the ratio between the number of stars and the mass of the dark matter halo is not constant. The reason for this is that the first supernovae in smaller halos can knock the rest of the gas out of those halos due to the extremely shallow gravitational potential presented by the dark matter in those halos. In fact, the smallest halo we expect to contain baryons is around $10^9 M_\odot$ [61] but the cut-off is gradual and one might expect outliers on both sides—smaller halos which do contain stars and larger halos which do not.

2.1. The Missing Satellites “Problem”

There are many historical claims that there is a lack of satellite galaxies observed in the local group relative to the number predicted in N-body simulations of dark matter [42–44]. However, the advent of more sensitive survey telescopes, such as the Dark Energy Survey, has shown the presence of many more of these small galaxies [62], while re-analysis of the problem has found different conclusions [63]. Anyway, it does not seem to be clear that there is a missing satellite problem. Nevertheless, warm dark matter suppresses the growth of structures below its free-streaming length (corresponding to the scale of smaller galaxies). We can therefore use the Milky Way satellites we observe to place constraints on the speed of dark matter particles in the early Universe—see, e.g., references [64–69]. Typically, the constraints obtained by these methods lead to constraints on the smallest objects being formed being less than $10^9 M_\odot$. This can be translated into a naive particle physics model by assuming that we are talking about some neutral particle which was at some point in thermal–chemical equilibrium with the primordial plasma but which froze out in the early Universe. When we do that, we typically obtain a constraint on its mass of being a few keV (see, e.g., reference [64]).

Recently, it has been claimed that in warm dark matter models, the situation is not as simple as it might seem. Since the smallest halos form close to the cut-off in the power spectrum, where some of the power has been erased through partially being eradicated through free streaming, those halos form later in the Universe and have a smaller concentration parameter due to their central density reflecting the density of the Universe at the epoch they formed. When they subsequently fall into larger conglomerations of dark matter, such as the Milky Way, their outer parts are more vulnerable to stripping, and larger halos end up looking like smaller halos [70]. It might be, therefore, that some of the constraints on the speed, and therefore the phase-space density, of dark matter in the early universe might need to be revisited in the future.

2.2. Lyman-α Forest Constraints

Warm dark matter suppresses the growth of structures at small scales, and this does result in smaller sub-halos not being formed and not observed today. However, it also of course affects the growth of structures at high redshifts. The Lyman-α forest is a series of absorption lines observed in the spectra of high-redshift quasars, which take place in gas clouds before the light reaches the Milky Way (see Figure 2). The lines are always at the 2–1 transition of hydrogen but because the clouds are at various different redshifts, they form a dense series of lines in the spectra, spread over a range of frequencies. The gas clouds in which these photons are absorbed are not anywhere near as dense as within galaxies such as the Milky Way such that gas pressure or energy loss is not relevant, and the gas density follows the dark matter density. By analysing the widths of these lines, observational astrophysicists probe the radial velocity dispersion of the gas and, through the virial theorem, the underlying density of dark matter.
Figure 2. Diagram of how Lyman-α absorption lines appear in quasar spectra as the light passes through clouds at different redshifts based on the simulations of reference [71]. Light from a quasar located at high redshift (high $z$) passes through clouds at various redshifts. Absorption within those clouds corresponding to the $1 \rightarrow 2$ transition in hydrogen and absorption features shows up at multiples of $1 + z$ of the original wavelength (Image reprinted from reference [71] according to creative commons re-usage policy).

Once these observations are turned into a power spectrum and compared with simulations, it is possible to work out which models of dark matter are ruled out. All observations of the Lyman-α forest appear to be compatible with the expectations of cold dark matter—see, e.g., references [72–75]. This places a very strong constraint on how warm dark matter could be.

It has also been shown that these constraints can also be applied to ultra-light dark matter, see for example reference [76] and the more recent analysis of data in reference [76].

2.3. Too Big to Fail

The too-big-to-fail problem [44,77,78] is basically the fact that we expect bigger sub-galaxies of the Milky Way than those that are observed. Some of the more massive halos we expect to be there are expected to be so massive that they are resistant to star-formation suppression, which occurs at re-ionization.

The problem could represent a challenge to the CDM paradigm, but one possible solution to the problem could be a mismatch in the central densities of galaxies, which are satellites of larger galaxies and those which are in the field. These central densities are smaller in simulations containing baryons than predicted by dark-matter-only simulations,
making it more likely that the baryons will exit upon star formation, hence them being “big” but “failing” anyway [79–85].

2.4. Other Probes of Small Scale Structures

It is still extremely important to find ways to measure the presence or absence of very small (<10^8 M_⊙) dark matter halos in galaxies. A constraint which has been pursued in some detail now is the idea that streams of stars will be perturbed by smaller dark matter sub-halos tidally disrupting their structures [86] and there are tentative indications of halos as small in mass as 10^7 M_⊙ being detected in perturbed streams around the Milky Way [87]. It is hoped that such analyses will be able to get down to the 10^6 M_⊙ scale in the future.

In the same way, there are new avenues to looking for a small-scale structure using strong lensing. For example, a recent study looked at magnification ratios in multiple images of the same object to look for a small-scale structure, and compared these with observations of the local distributions of dwarf galaxies, to come to a constraint that shows that halos with masses around 10^7 M_⊙ still exist [88].

3. Cores vs. Cusps

N-body simulations of dark matter with initial random density perturbations suggest that they should form density halos with steep profiles—the density of dark matter should rise steadily towards the centre of the halos. Typically, this is modelled by an NFW profile where the central density rises, such as \( \rho \propto r^{-1} \) as \( r \to 0 \), but the authors responsible for the acronym NFW themselves suggest that an Einsato profile is more suitable [89]. Nevertheless, it is clear that there is indeed a rising at the middle of simulations, which is referred to as a cusp, rather than a large constant central density, referred to as a core.

It has been claimed that this is not what is observed in nature—i.e., it has been claimed that the dark matter in actual galaxies does not follow this behaviour at their centre and is cored [45–49]. In order to work out the distribution of mass in a galaxy which includes the mass which cannot be seen, such as dark matter, one needs to look at the kinematics of tracers, such as stars or gas. For a well-behaved spiral galaxy, this would involve looking at the rotation curves of stars in the disc and gas in the outer areas, as well possibly as looking at the velocity dispersion of stars in the central bulge, which is more spherical and often with a less well defined coherent rotation. For a Galaxy such as the Milky Way, the baryonic matter is thought to dominate the gravitational potential quite far down into the central region and it is challenging to obtain rotation curves in that area [90].

3.1. Measuring Cores vs. Cusps in Dwarf Galaxies

Ultra-faint dwarf galaxies are in principle brilliant places to start to look for dark matter. There are very few stars and lots of dark matter so the mass to light ratio is very large. We deduce the presence of the dark matter by observing the velocity dispersion of the stars—this is done by looking at their spectra and using this to deduce their line-of-sight velocity. Unfortunately, these objects are too far away (typically 50–200 kpc) to deduce any component of their proper motion, so we have to make deductions about the gravitational potential simply based upon the line-of-sight velocity readings.

The Jeans Equation [91] arises by taking moments of the collisionless Boltzmann equation while assuming a spherically symmetric steady-state solution. The equation takes the form

\[
\frac{1}{\nu} \frac{\partial}{\partial r} \left( \nu \sigma_t^2 \right) + 2 \frac{\beta \sigma_t^2}{r} = -\frac{GM(r)}{r^2},
\]

where \( \nu(r) \) is the density of tracers (the stars to be observed, not the objects responsible for the gravitational potential) and \( \beta(r) \) is the velocity anisotropy:

\[
\beta = 1 - \frac{\sigma_t^2}{\sigma_r^2}.
\]
Here, $\sigma_t$ and $\sigma_r$ are the tangential and radial velocity dispersions of the tracer stars, respectively.

Solutions to the Jeans Equation (1) are subsequently used to determine the line-of-sight velocity dispersion, given by reference [92]:

$$\sigma_{los}^2(R) = \frac{2}{\Sigma(R)} \int_R^\infty \left(1 - \frac{\beta R^2}{r^2}\right) \nu \sigma_r^2 \frac{rdr}{\sqrt{r^2 - R^2}},$$

where $\Sigma(R)$ is the surface-mass density of tracers (usually stars) at projected radius $R$ (3D density flattened onto a plane as 2D surface density). In order to reconstruct the density of dark matter in the inner regions of a dwarf galaxy, one needs to work from observations of $\sigma_{los}(R)$ to attempt to find the mass $M(r)$.

The relationship between the stellar velocity dispersion and the gravitational potential therefore depends critically upon the velocity anisotropy parameter $\beta$. Since we only obtain the projected line-of-sight velocities of the stars, we need to marginalise over this nuisance parameter—there is a degeneracy between the cuspiness of the profile and the velocity anisotropy parameter.

It is well known [93] that the Jeans equation only gives definitive constraints on a dwarf spheroidal galaxy’s total mass at one particular radius, namely where the logarithmic slope of the stellar density profile $d \ln \rho(r)/d \ln r = -3$. This point often coincides with the half-light radius—the projected radius on the sky within which half of the stars lie. At all other radii, the true mass is masked to some extent by the degeneracy with different assumptions for the anisotropy $\beta$.

One of the most interesting claims for the cored nature of these dwarf galaxies came about from the identification of multiple populations of stars in dwarf galaxies with different metallicities. By treating these different populations as separate tracers, it was pointed out that one could obtain a measure of the mass at the half-light radius for each of these populations independently [48]. By having two mass measurements at different radii, it is then possible to deduce the logarithmic slope of the density profile and it has been claimed using this method that many dwarf galaxies may possess cores [48]. It turns out, however, that this method is not free from systematics, and applying it to stellar populations inside simulations shows that the gradient of the density profile can end up being affected by observational perspective effects (i.e., it depends what direction you are looking from) [94].

3.2. Using Kurtosis to Solve the Problem

To try to see if a more robust treatment of the velocity dispersions might be useful, we looked at including information from higher moments of the velocity dispersion, in particular, not only the second moment of the velocity dispersion (the variance) but also the fourth moment, and their ratio, giving us the kurtosis [95,96]. We found that this extra information could help to cancel the degeneracy between the density profile and the anisotropy parameter. In particular, in the case of the Sculptor Dwarf, we found that the dark matter distribution was cuspy using these methods, whereas the method using multiple populations showed it to be cored.

Further advances in this area were made when the higher-order kurtosis methods were applied more generally using new data, finding that some dwarfs were cored and some were cuspy. Interestingly, it seems that there is a connection between cored dark matter profiles and a recent history of star formation in dwarf galaxies [97,98]. This could occur when non-adiabatic changes in the gravitational potential due to shockwaves created during bursts of star formation effectively heat up the dark matter, releasing the dark matter in the core.

This makes it increasingly difficult to ascribe the cores inside dwarf galaxies, should they appear there, to be due to particle physics, but rather they seem to be due to astrophysics. In particular, it means that some analyses which have used the idea that dwarf galaxies are cored may be less reliable than they were previously [99].
3.3. Cores vs. Cusps in Larger Galaxies

On larger scales, the situation is more complicated. In particular, in the region where galaxies and their dark matter halos appear to weigh more than $10^{10} M_\odot$ but significantly less than the $10^{12} M_\odot$ corresponding to the Milky Way, there are a population of galaxies with cored dark matter profiles [49]. While this can be explained by star formation rates inside those galaxies, and the latest simulations effectively model it, it is not something which is universally understood.

3.4. Dark Matter Simulations Containing Baryons

Computer simulations of galaxies involve a combination of gravity-only N-body simulations for the dark matter with smooth particle hydrodynamics (SPH), which are supposed to model the gas and the stars. While the gravity-only part is less controversial, the SPH side has many free parameters which represent sub-grid physics—physical processes which occur on length scales smaller than what can actually be modelled (for example, individual stars and their surrounding gas). There are a variety of free parameters, such as star formation rate and viscosity, etc., which are tweaked to get galaxies as close to what is involved in nature as possible [100].

It is clear that in order to explain many pieces of data, it is necessary for the dark matter to react to the behaviour of baryons in certain galaxies—it is not enough for dark matter to keep the same density profiles as what comes out of N-body simulations without baryons in. For the effect of adiabatic contraction see reference [101], and an adiabatic gas shocking (see reference [82]) means that the baryons push and pull the dark matter around. The way that the dark matter reacts to baryons in this way is itself part of the reason why non-minimal models of dark matter or even modifications of gravity, such as MOND, have been so popular. There will always be criticisms that the way parameters are chosen in SPH simulations is arbitrary and only serves to model what is observed in reality. Of course, this is not a source of shame for the modellers, it is simply all they can do, since the parameters are meant to model extremely complex behaviour at scales far below the resolution of the simulation. Until computers can model individual solar systems in galaxies hosting different stars and gasses (which will never happen), this will always be an inexact science, and this does not reflect a weakness of dark matter numerical modelling—it is simply a feature of the field. A lot of research has gone into understanding these sub-grid physics issues [102].

4. Phase-Space Constraints on Fermionic Dark Matter

There is plenty of motivation for relatively light fermionic dark matter candidates, see, for example, reference [19] for a review. Sterile neutrinos are very commonly invoked in particle physics to solve anomalies at neutrino experiments and are very common in particle models which aim to explain neutrino masses and the asymmetry between matter and anti-matter, see, for example, reference [103]. Critically, keV-scale sterile neutrinos can be produced in the early Universe in the right abundance to explain the dark matter we see in the Universe today, e.g., references [52,104].

The most obvious test of sterile neutrinos is related to the fact that such scenarios would in general imply that the dark matter would be slightly warm because they typically possess some residual kinetic energy at the epoch of galaxy formation. They would then be sensitive to the constraints on structure formation outlined in the previous section. X-ray constraints further constrain their parameter space, since sterile neutrinos can decay into active neutrinos, and the region of parameter space sensitive to such warm dark matter masses corresponds to X-ray energies, e.g., see references [105,106] for a review.

More generally, if we are completely agnostic about the nature of dark matter, then there is a reasonably good chance it is a fermion (not forgetting the possibility that dark matter is made up of primordial black holes), so it is interesting to find out if we can place constraints on such scenarios using the fermionic nature of the particles and the way this feature will alter the phase space of the dark matter in halos.
As first envisaged by reference [107], there is a limit to the extent to which fermionic dark matter particles can be compressed within a halo—see, e.g., references [108,109]. This class of constraints applies to all light fermionic dark matter particles, whether they are warm dark matter or not.

In our work, on this topic [110], we used the conservation of phase-space density via Louville's theorem on dark matter in dwarf spheroidals in order to put a lower bound on the fermionic dark matter mass. On the data side, we benefit from the improved kinematical analysis of dwarf spheroidal galaxies from reference [97] explained in the earlier section, which allows us to better determine the density profile. We obtain more constrained spreads for the density of dark matter as a function of radius using Jeans techniques with higher-order moments of the line-of-sight velocity. We then solve the Jeans equation for the dark matter using priors on the distribution of \( \beta(r) \) from warm dark matter simulations.

We are then able to constrain the phase space, i.e., the density of dark matter in position and velocity space. Since Louville's theorem states that this density is conserved, we can then compare this constraint to three different situations which represent the initial phase space of the fermionic dark matter in the early Universe (see Figure 3).

**Figure 3.** Bounds on the mass of fermionic dark matter for the full set of dwarfs considered in the work [110]. On the left, these bounds are computed using the Pauli exclusion principle, in the middle we use phase-space arguments for relativistically decoupled thermal fermions, and on the right, we do the same with non-resonantly produced sterile neutrinos. \( 1 \sigma \) and \( 2 \sigma \) constraints are denoted by the darker and lighter colours. The dark vertical line inside the \( 1 \sigma \) constraint is the central value. For each dwarf, the bounds above the dotted lines are obtained using maximal coarse-graining, while those below are obtained using Gaussian (this is taken from the paper [110] in accordance with the usage permissions of Oxford University Press who retain the copyright).

5. Self-Interacting Dark Matter

Self-interacting dark matter refers to models where dark matter particles can interact with each other and can therefore have effective pressure [111]. We know that dark matter cannot have long-range interactions with very light mediators, since if that was the case, it would lose energy by emitting those mediators as real particles and fall into the centre of galaxies. It is possible, however, for dark matter to have relatively high short-range self-interactions. In fact, if the dark matter had a mass roughly equivalent to a neutron, then its self-interaction cross-section could be as large as the neutron–neutron cross-section...
mediated by the strong force, corresponding to a cross-section per mass of roughly a \( \text{cm}^2 \text{ per gram} \) \cite{112–117}.

This constraint on the cross-section is obtained by looking at dwarf galaxies around the Milky Way \cite{118–122} and also by studying galaxy clusters \cite{123–125}.

Typically, dark matter halos made out of self-interacting dark matter will exhibit cores at the centre rather than cusps, as the central regions start to exhibit a hydrostatic equilibrium \cite{59,126–129}. Dark matter self-interactions can also wipe out “rugby ball” shapes in dark matter halos, which are typical in non-interacting cold dark matter scenarios \cite{101,125}.

Recently, it has been shown that self-interaction particle physics models that would result in many interactions with small exchanges of momentum cannot be mapped directly onto models with fewer interactions and larger exchanges of momentum \cite{130}. The fact that there are quantitative differences between the results of these simulations to more traditional self-interacting dark matter simulations show that the subtle effects of self-interaction due to more complicated interaction physics require careful analysis.

6. The Diversity Problem

Potentially strongly related to the previous section, the divisity problem \cite{131,132} is a statement that the rotation curves of galaxies in reality are a lot more varied than the rotation curves of simulated galaxies observed in cold dark matter simulations with SPH modelling baryons.

The problem is that the rotation curves of the galaxies which grow in such simulations are a lot more uniform than the rotation curves which are observed in real galaxies in nature. This sets off alarm bells, since a lot of the explanations of apparent discrepancies between simulations and observations (for example cores) are based upon SPH arguments.

One explanation is that the dark matter could be self-interacting. With self-interacting dark matter in baryonic simulations, the density of DM depends more strongly on the baryonic potential \cite{133–135}. With the Dark matter reacting more sensitively to the non-linear baryonic physics, the possibility of diversity increases \cite{136–141}.

7. Planes of Satellite Galaxies

The planes-of-satellites phenomenon refers to the fact that there is some evidence for a favoured plane of the orbits of satellite galaxies around the Milky Way, the Andromeda Galaxy, and Centaurus A \cite{142–150}. This seems difficult to understand in cold dark matter, since the collapse/formation of galaxies in those models from random initial conditions seem to result in smaller galaxies joining randomly from different directions with uncorrelated angular momentum vectors. Different mechanisms such as the accretion of dwarfs at the same time, the arrival of dwarf galaxies from filamentary structures, and the tidal disruption breakup of large-scale co-rotating structures have been considered but do not seem to explain the features adequately \cite{151}.

At the current time, it is not clear how ubiquitous this phenomenon is, nor what the probability of the observed correlations in orbits are. It is clear that this might be a challenge to the typical cold dark matter scenarios, but it remains to be seen to what degree. It seems quite difficult to solve this problem with dark matter.

The suggestion that the co-rotating dwarf galaxies forming the planes could form from tidal features pulled off the Milky Way, which then collapse, is tempting to imagine, but suffers from two obvious problems, the first is that such tidal features would only be made out of baryons, since dark matter has too much energy/entropy to form the elongated tidal features suitable, and yet the dwarf galaxies observed around the Milky Way all exhibit internal dynamics, suggesting the presence of dark matter. This could be solved with a theory such as modified Newtonian dynamics (MOND) \cite{152}. The second problem is that if these dwarf galaxies did form from matter coming from the Milky Way, then the stars in them should have the same metallicity as the host, while the metallicity of the stars in dwarfs around the Milky Way is consistent with them being primordial in origin.
A close encounter in the past between Andromeda and the Milky Way could have explained the presence of these galaxies, however, in CDM, this would also have been extremely disruptive for both large galaxies, and dynamical friction would have resulted in their coalescing. It has been claimed in MOND, however, that this would not be the case, since only the stellar dynamical friction would be present and the galaxies could survive an encounter.

8. MOND and the Radial Acceleration Relation

Modified Newtonian dynamics, or MOND, is an attempt to explain the behaviour of galaxies, assuming that the low acceleration limit of Newtonian gravity (and hence general relativity, for which Newton’s theory is of course a non-relativistic approximation) is modified, and there is a lower limit on the acceleration induced by massive bodies [152–154]. MOND can (arguably) solve a variety of problems at the galactic scale in a more (arguably) simple way than dark matter. However, it has problems, especially on other scales, typically larger scales, such as cluster scales and cosmological scales, as we will describe below.

8.1. Successes of MOND

MOND can in principle explain a variety of behaviours without dark matter. One observation that we have not mentioned so far is the Tully–Fisher relationship, which is an observed relationship between the rotation and the luminosity (and therefore presumably the baryonic mass) of galaxies [155]. At first glance, it appears that this relationship has no possible explanation in cold dark matter, although it has been shown that this relationship could arise from baryonic feedback on dark matter halos [156]. This is not, however, the only relationship which can be explained with MOND—most obviously, it can naturally lead to flat rotation curves without the need for a dark matter halo with a region where the density gradient drops like $\rho(r) \propto r^{-2}$.

MOND can also explain the plane of satellite galaxies, since it is possible to imagine scenarios where close passes of large galaxies in the past occurred such that satellites could have been ripped from one to the other without the tidal disruption of those galaxies being so extreme that they end up coalescing [157].

MOND can also explain the apparent cores inside the “larger” “dwarf” galaxies—the low-surface-brightness galaxies which lie between the super-small ultra-faint dwarfs and the big galaxies around the Milky Way. In MOND, the fact that these galaxies are low surface brightness and the fact that they do not exhibit the rapid rotation one might expect from cold dark matter is naturally accommodated; in MOND, the the only source of gravitational potential is baryons, so low surface brightness means low baryonic mass, which means apparently cored dark matter halo, should one choose to interpret the rotation as dark matter [158].

Another success of MOND is the ability of the theory to explain features in the rotation curves that are related to features in the baryonic density—in the regions of the galaxy where the dark matter gravitational potential dominates, one expects variations in that potential due to baryons to be insignificant, however there are often fluctuations in the rotation curves associated with those baryonic density changes.

A large data set of rotational velocities drawn at different radii in different galaxies with different sizes has been produced and is called the SPARC database [159]. The relationship between the expected acceleration due to the observed baryonic distribution (assuming Newtonian gravity) and the actual observed acceleration of stars (using observation of the stars) is plotted, and there is a remarkable relationship between the two across all these different scales and galaxies, which is sometimes referred to as the radial acceleration relation [160]. It should be noted that this relationship is not entirely consistent with a particular acceleration scale but rather a unified relationship between the expected and observed acceleration. It has been claimed that this can arise naturally in cold dark matter with baryons using baryonic feedback.
8.2. Problems with MOND

Despite solving a myriad of, although not all, problems on galactic scales, MOND is very bad at explaining the behaviour of galaxy clusters, where it is essential to have some component of dark matter.

Most notable is the failure of MOND to explain the Bullet Cluster. This object seems to be two galaxy clusters which have collided and passed through each other [161]. The majority of baryons in galaxy clusters are not in the galaxies themselves but in the hot intergalactic intra-cluster gas (the gas between the galaxies but inside the cluster), which can be studied via its X-ray Bremsstrahlung emission. An observation of this gas in the case of the Bullet Cluster shows clearly that the smaller cluster has punched through the larger cluster, leaving a shockwave reminiscent of that which a bullet would make on striking some solid material, hence the name. The two bodies of hot gas, one from each cluster, seem to have been stuck in the middle of the collision due to their mutual pressure, while the galaxies inside the two clusters have continued without such impedance and are no longer in the same place as the gas. Most of the baryonic matter is stuck in the middle of the collision, while the galaxies themselves have sailed away through the collision and out onto the other side in both directions.

How is this a constraint on MOND? Well, gravitational lensing studies of the cluster show that the majority of the mass lies in the same place as the galaxies, not in the same place as the gas, which got stuck in the middle [6]. Since we know most of the baryons are in the gas, this tells us that there is a big component of mass which we cannot see, presumably the dark matter. It is very difficult to explain the Bullet Cluster using MOND. There are criticisms of the Bullet Cluster—some people say that there are very few really reliably observed objects.

Perhaps equally importantly, MOND fails to explain the observations in the CMB. This is less obvious to prove because the kind of phenomenological tweaking of the non-relativistic relationship between the gravitational potential and the acceleration at the heart of the MOND paradigm is clearly not satisfactory to predict the behaviour in the early with-density modes entering and leaving the horizon in both matter-dominated and radiation-dominated epochs. To track the evolution of such density modes requires a fully relativistic theory, such as general relativity—its Newtonian limit will not suffice.

However, the basic theory of MOND has no UV completion, it is purely phenomenological. It has, however, been possible to write down a fully covariant theory which leads to MOND-like behaviour, TeVeS [162]. This theory is thought to be the minimum kind of modification of gravity which can give rise to MOND-like behaviour, which would explain galaxies but also arise from a fully relativistic covariant Lagrangian. Critically, it is well defined enough to be able to perform calculations related to the growth of perturbations in the early Universe. Unfortunately, it has been shown that while the first, second, and even third peaks of the CMB TT spectrum can be reproduced without too much discrepancy, higher peaks cannot be [163].

8.3. Particle Attempts at MOND

There have been a few attempts to model MOND using a particle theory. One important notable attempt is the idea of dark matter superfluidity—that in smaller halos, such as galaxies, the dark matter becomes a superfluid condensate and has a characteristic behaviour, which automatically leads to flat rotation curves and can be related to the micro-physics of the dark matter [164,165]. The model also has an interesting feature, namely that in clusters, there is no MOND-like behaviour. The explanation for this is that the increased velocity dispersion in clusters means that the temperature of DM is sufficiently high to prevent the superfluid from forming.

Another interesting attempt to combine dark matter and MOND was made in reference [166], where the authors showed that a scalar field non-minimally coupled to matter with non-standard kinetic terms can under certain limits lead to MOND-like behaviour, but also at other times lead to a condensate which acts like dark matter. It is remarkable
that this can be achieved, but shows that in the examples where MOND fails, dark matter is really required in one way or another to pick up the slack. Having both MOND and dark matter having their origin in the same physics may be seen as a positive to varying degrees, depending upon the outlook of the reader.

8.4. Unscientific Thoughts and Observations

There are very specific, difficult, possibly intractable problems with MOND, as we have seen.

There is an unusual dichotomy in the field when it comes to the subject of MOND and theories similar to MOND—if one goes to a particle physics conference there will typically be a significant fraction of the time devoted to models of particle dark matter, but anyone presenting on the subject of MOND will, in my personal experience, face a barrage of scepticism. There are, of course, a good selection of reasons for this, since at a particle physics conference, the physicists naturally subconsciously expect dark matter to be explained as a particle. Furthermore, particle physics model-building phenomenologists notoriously work to “satisfy all the constraints”, so if MOND cannot cope with cosmology and clusters, such physicists simply do not care what it can do on galactic scales.

I would say that, as alluded to the in previous section, the younger generation of scientists are more open to theories that have limits which act like MOND and are interested in explaining how galaxies can exhibit MOND-like behaviour in other ways which do not necessarily involve modifications to gravity [165] and which do not mess up cosmological observations.

On the other hand, in a room full of galactic astrophysicists, those in the room more open to MOND are (depending on the conference, again in my own personal experience) taken very seriously and are viewed as being much more mainstream scientists. Perhaps a lot of this is to do with the fact that those closer to them understand the difficulty they faced in obtaining and/or analysing the data they are using to make their conclusions.

It is, in fact, very easy to argue in favour of MOND. When we look at the history of the discovery of gravity, it was Kepler’s observations of astrophysical objects (the planets) which led to Newton’s development of the theory of gravity. Why then, when we look at galaxies and we see them rotating in an apparently anomalous way in their outer regions, do we not assume that our theory of gravity is wrong?

History has certainly played a role in this. The prediction and ultimate discovery of neutrinos and the knowledge that most of the photons in the Universe are in the CMB, which had not been seen until the 1960s, both perhaps make it easier to understand that some particle species so far undetected could be around. Furthermore, the fact that physics beyond the standard model was expected, naturally led independently to theories containing dark matter candidates.

The need to explain the hierarchy between the electroweak scale and the Planck scale led to theories containing new particles and forces around the same scale, which naturally led to WIMP candidates. More specifically, supersymmetry very nicely cleaned up the problem of the electroweak hierarchy but led to proton decay without the inclusion of an additional symmetry R-parity, which also rendered the lightest supersymmetric particle stable and a natural dark matter candidate [167]. It should be noted that, regrettfully, at the time of writing, there is no evidence at colliders for either supersymmetry nor for dark matter candidates or their mediators at the electroweak scale.

Similarly the Strong CP problem (the lack of CP violation in QCD, as demonstrated by the vanishingly small electric dipole moment of the neutron) led to the concept of the axion, which is itself another dark matter candidate [18].

So, these very strong particle physics motivations may have biased the particle physics community into believing that dark matter naturally has to have a particle physics explanation. I think it is important to have an open mind about these things and not only to understand the failures of MOND but also its apparent successes, and to look at both critically.
However, having said that, and having examined the evidence carefully, MOND ultimately is a fascinating theory which explains many mysterious features of galaxies but simply does not work at larger scales. It does not seem appropriate to suggest the theory is as good at explaining the Universe we live in as the theory of dark matter.

9. Conclusions

In this brief review, I have tried to give an update from the point of view of an enthusiastic amateur as to the status of various observations from astronomy, which might help shed light on the nature of dark matter.

We know that dark matter is not very good at interacting with us or itself, otherwise it would not do what we need it to do. All of these astrophysical tests rely upon dark matter almost interacting with itself strongly or just having enough kinetic energy for its behaviour to be distinguishable from cold dark matter, but it is clear that cold dark matter without any self-interaction is a strong paradigm.

It is not at all clear that there is any modification of ΛCDM required in order to explain the observations of galaxies. Some scientists claim that there are definitive deviations, but there does not seem to be, to my knowledge, any discrepancy from vanilla ΛCDM, which is robust and universally agreed upon. What this means is that if there are new effects, such as warm dark matter or self-interacting dark matter, which will be observable in the coming decades, then they must be quite finely tuned to be very close to what has already been ruled out or they will never be observed. For example, if dark matter is slightly warm but colder than what would have been required to explain the missing satellites of the Milky Way (before they turned up), then it would only suppress those satellites less than $10^7 \, M_\odot$, which do not contain any stars anyway and therefore cannot be observed.

Self-interacting dark matter can only have an effect on galactic scale objects if its cross-section is close to the ~cm$^2$ g$^{-1}$ level which is currently constrained, since that cross-section, combined with the size and density of galaxies, is what is required to get about one interaction per crossing. Anything less will make little difference to the galactic structure; however, there are surely many particle physics models which could predict much smaller self-interaction cross-sections.

Because of this, any way in which we can learn about the nature of dark matter using astrophysical anomalies will by necessity be rather fine tuned, and we will have to be lucky to observe such behaviours. That is not to say that we should not try to constrain such deviations from ΛCDM as much as possible.

Researchers doing simulations of gravity have made tremendous progress in recent decades, and the very fact that the galaxies which emerge from their codes are so similar to what we observe is truly remarkable. Cynics would note that the codes are built to do this, so it is not surprising. As they understand more about the physics of sub-grid physics, simulators will presumably continue to refine their trade and make more and more independent predictions and tests, which will constrain their SPH code.

New telescopes will help the field greatly. The Vera Rubin Observatory will provide a huge amount of information about dwarfs and also streams of stars, as well as potentially looking for lensing events due to low-mass dark objects. This information should help physicists, if not find out more about what dark matter is, then certainly more about what it is not.

Funding: The author is supported by the European Research Council under the European Union’s Horizon 2020 Program (ERC Grant Agreement No 648680 DARKHORIZONS) and by the STFC Grant ST/P000258/1.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.
Acknowledgments: I am grateful for references suggested by Stacey McGaugh, Marcel Pawloski and Michael Boylan-Kolchin

Conflicts of Interest: The author declares no conflict of interest.

References


