Why did the dark matter hypothesis supersede modified gravity?

Antonis Antoniou

Lichtenberg group for the History and Philosophy of Physics
University of Bonn

UNTERSTÜTZT VON / SUPPORTED BY
Alexander von HUMBOLDT STIFTUNG

Setting up the problem

Dark matter related anomalies can, in principle, be explained by making the following two working hypotheses:

WH1: Dark matter

The postulation of a non-baryonic form of matter which interacts with baryonic matter only via gravity.

WH2: Modified Gravity

The modification of standard Newtonian Dynamics as a limit of a relativistic theory.

Question

Given that relevant observations can, in principle, be explained either by postulating dark matter or by modifying gravity, what motivated the pursuit of each of these hypotheses in the 1980s?

Motivation

1. Back in the 1980s, physicists were not making a choice between two complete theories but rather, a choice between working hypotheses. Why did the DM hypothesis supersede compared to the MG hypothesis?
2. A similar situation of underdetermination arises in the Dark Energy case, with the difference that the MG hypothesis is not excluded. Understanding the dark matter case also sheds light on the reasons for this difference.

Aim

To examine the scientific landscape of the 1980s, and show why the pursuit of WH2 (Modifying Gravity) was a much more challenging and perhaps unmotivated endeavour, compared to the pursuit of WH1 (Dark Matter) and its eventual integration in the ΛCDM.

Outline

Missing mass in clusters

- **Zwicky (1933)**: Radial velocity measurements in Coma cluster using the virial theorem: galaxies moving too fast / the cluster appeared to contain several hundred times more mass than the visible mass
- Did not receive much attention / Discrepancy attributed to observational errors in accounting mass and light in galaxies
- **Schwarzchild (1954); Van de Hulst et al. (1957); Oort (1960);** : re-examine mass-to-light ratio in galaxies / conclude that M/L is significantly higher than what is expected for normal stellar populations / still doubts... discrepancy attributed to undetected baryonic mass, e.g. molecular gas

Extended Rotation curves

- Developments in radio astronomy in 1970s provide the opportunity to probe rotation curves of neutral gas well beyond the optical image of the galaxy, esp. via 21 cm observations
- **Rogstad & Shostak (1972)**: find that rotational velocities in five spiral galaxies rise sharply to a maximum value and then remain flat and "confirm the requirement for low-luminosity material in the outer regions of these galaxies."
- **Rubin, Ford & Thonnard (1980):** precise spectroscopic observations of the rotation curves of spiral galaxies using visible emission lines of hydrogen and nitrogen in 21 spiral galaxies / all rotation curves appear to be flat

Stability of spiral galaxies

- **Hohl (1971):** studys the dynamics of galactic systems in N-body simulations / finds that cold rotationally supported disks of stars are unstable, and develop into pressure supported systems / But our galaxy doesn't seem to be pressure supported
- **Ostriker & Peebles (1973)**: Show that the presence of a massive halo in the outer regions of spiral galaxies provides the necessary stability / speculate that the halo consists of very low mass stars or white dwarfs (i.e. objects with low M/L) / suggest further observational searches 'to see if numerous faint high-velocity stars exist in the solar neighbourhood'.

Structure formation and CMB anisotropies

- **Peebles (1965, 1966, 1968)**: Highlights the structure formation problem: in order to produce the observed structure of the Universe, small density fluctuations in photon-baryon fluid must occur in the early Universe which would correspond to comparable fluctuations in the temperature of CMB / such fluctuations were not observed
- **Gunn et al. (1978):** *any* heavy stable particle that is a cosmological relic and does not couple to photons can solve the structure formation problem, insofar as it dominates the matter budget of the universe
- **Peebles (1982), Vittorio & Silk (1984), Bond, Efstathiou & Szalay (1983; 1984)**: propose dark matter as a solution to the problem of structure formation: if there is a non-interacting dark matter fluid with $\Omega_{dm} \approx 1$, fluctuations in it can begin to collapse when matter becomes gravitationally dominant, i.e. much earlier than the decoupling epoch / consistent with CMB patterns of temperature anisotropies at the time

The state of the art in the 1980's - Summary

Four main problems / anomalies:

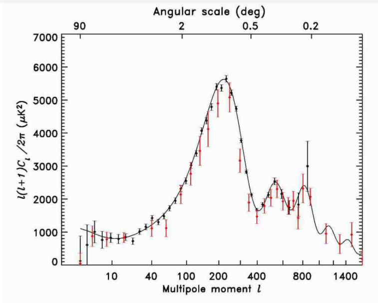
1. Missing mass in galaxies / high velocity dispersions
2. Flat rotation curves
3. Instability of spiral galaxies in simulations
4. Structure formation and CMB anisotropies

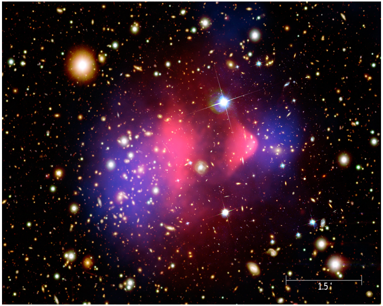
And two further widely accepted facts:

5. A strong preference for a closed universe
6. The renaissance and establishment of general relativity

Justification – The coup de grâce

- COBE (1992) and later WMAP (2003) find the missing primordial fluctuations confirming the necessary role of DM in large-scale structure formation
- 2006 - Gravitational lensing - The Bullet Cluster





Laudan (1978) on pursuing a working hypothesis

Three stages:

- **Discovery**: The process by which a theory or hypothesis is generated and proposed
- **Pursuit**: The further investigation of a possible hypothesis
- **Justification**: The decision process by which a hypothesis (or theory) is being accepted (or rejected).

Question

What reasons and motivations guide the scientific community to pursue a hypothesis and eventually embed it in a complete theory?

Four criteria

1. Problem-solving potential
2. Compatibility with established theories
3. Feasibility of incorporation
4. Independent testability

Problem-solving potential

Problem-solving potential

The potential of a working hypothesis to address different types of data/observations and solve a multitude of problems at once.

- **DM**: Solves, in principle, all four problems: missing mass, rotation curves, stability of spiral galaxies, structure formation. It also aligns with the preference for a closed universe.
- **MG**: Solves only dynamically related problems: missing mass, rotation curves, and stability of galaxies. Says nothing about the structure formation problem and CMB patterns. Incompatible with a closed universe.

Compatibility with established theories

Compatibility with established theories

The compatibility of a working hypothesis with established scientific knowledge and various widely-accepted scientific principles at a given time.

- **DM**: Fully compatible with the general theory of relativity and the Big Bang model of the universe.
- **MG**: Requires the modification of a well-established and rigorously tested physical theory on the weak-field limit. Milgrom's initial formulation violated basic principles such as conservation of momentum and equivalence principle.

Feasibility of incorporation

Feasibility of incorporation

The feasibility of incorporating a working hypothesis into existing or new scientific theories.

- **DM**: Does not require any modification of GR. / Further motivated by the development of supersymmetry in the 1970s which had already provided candidate particles for dark matter (LSP) in addition to the already existing (non-baryonic) neutrino.
- **MG**: Requires the modification of the phenomenology of Newtonian dynamics, and consequently of general relativity in the galactic scales in which GR was strongly tested. Building a new theory of gravity that predicts dark matter related phenomena while at the same time maintains the experimental success of GR and NG in galactic/solar scales was – and still is – an extremely difficult task.

Independent testability

Independent testability

The, in principle and in practice, possibility of testing the validity of a working hypothesis outside the domain for which it was initially proposed.

- **DM**: The dark matter hypothesis can be tested in a number of different ways: direct searches, indirect searches, collider searches, gravitational effects on large scales, precision measurements of cosmological observables (CMB, SN Ia, BAO).
- **MG**: The testability of a modified gravity hypothesis depends on the details of the final theory in which it will be incorporated. In the case of MOND, it was not clear at the time how to test the theory outside the domain of galactic dynamics (cf. Milgrom (1983): 'At the moment I cannot suggest a feasible laboratory experiment to test the ideas discussed above.')

Modifying gravity for Dark Energy

Under the assumption that the cosmological constant, Λ, vanishes, dark energy phenomenology can be equally explained by two hypotheses:

- **WH1:** To postulate a modified form of matter with negative pressure (e.g. quintessence, k-essence, perfect fluid etc.)
- **WH2:** To modify standard gravitational dynamics in a way that generates the accelerated expansion of the universe without the requirement of a cosmological constant or an exotic type of matter (scalar-tensor gravity, $f(R)$ gravity, braneworld models etc.)

But...

Unlike the dark matter case, the scientific community is much more tolerant to the possible modification of gravity as a possible explanation of dark energy phenomenology. Why?

Four reasons

1. Feasibility of incorporation: the two competing hypotheses are integrated in GR by modifying the rhs (matter) and the lhs (gravity) of the EFE respectively. From the point of view of GR, dark energy as modified matter has no fundamental physical difference from dark energy as modified gravity. No direct effects on small scales.
2. Problem solving potential: Both hypotheses can equally solve dark energy related problems (accelerating universe, age problem, critical density etc.)
3. Classical modifications of GR were already pursued before the observation of an accelerating universe via SN Ia measurements in 1998 for various reasons (e.g. for the validation of Mach's principle, due to the non-renormalizability of GR etc.)
4. Dark energy as modified matter faces its own problems which keep the hypothesis of DE as MG 'alive' (very low field mass compared to other fields, unexplained absence of coupling to other fields etc.)

Conclusions

- The dark matter and modified gravity hypotheses emerged as a result of a number of observations and theoretical developments.
- The prevalence of the pursuit of the DM hypothesis after the early 1980s can be jointly justified by the problem-solving potential, the compatibility, the feasibility of incorporation and the independent testability of the competing hypotheses.
- A similar situation arises in the DE case. But, unlike the dark matter case, the feasibility of incorporation and the problem solving potential of the two hypotheses is rather equivalent, which – partly – explains why dark energy as modified gravity was further pursued after 1998, and is still considered as a plausible hypothesis.

Thank you.

Sources / Selected bibliography:

- Bertone, G., & Hooper, D. (2018). History of dark matter. *Reviews of Modern Physics*, 90(4), 045002.
- De Swart, J. G., Bertone, G., & van Dongen, J. (2017). How dark matter came to matter. *Nature Astronomy*, 1(3), 0059.
- Frankin, A. (1993). Discovery, pursuit and justification, *Perspectives in Science*, 1(2), 252-284.
- Laudan, L. (1978). *Progress and its problems: Towards a theory of scientific growth*. University of California Press.
- Milgrom, M. (1983). A modification of the Newtonian Dynamics as a possible alternative to the hidden mass hypothesis. *The Astrophysical Journal*, 270, 365-370.
- Sanders, R. H. (2010). *The dark matter problem: a historical perspective*. Cambridge University Press.
- Peebles, P.J.E. (2020). *Cosmology's century: An inside history of our modern understanding of the Universe*. Princeton University Press.

Thank you.


THEY ALL ASK "WHAT IS DARK MATTER?" AND "WHERE IS DARK MATTER?", BUT NOBODY ASKS "HOW IS DARK MATTER?"
TOM GAULD for NEW SCIENTIST