Large-scale Asymmetry of Galaxy Spin Directions - A Comparison of 12 Datasets

Lior Shamir

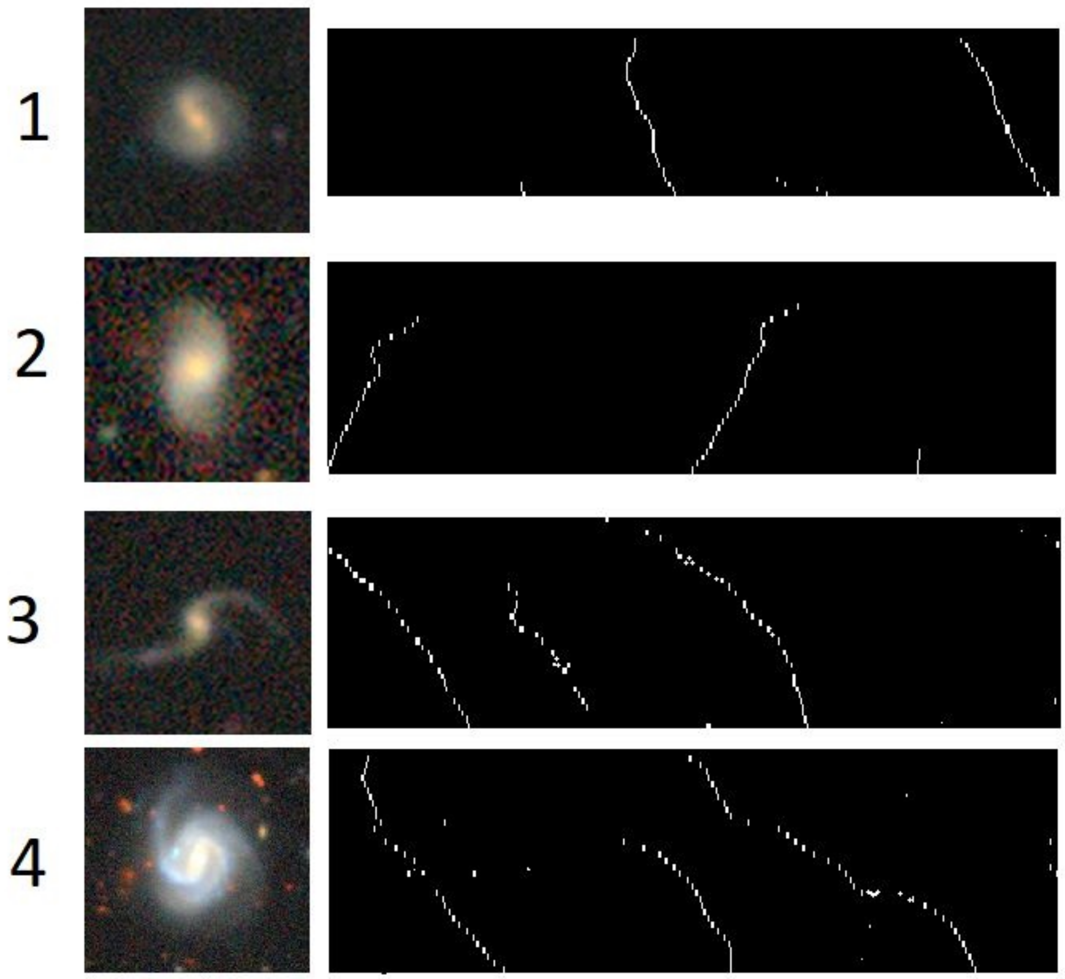
Kansas State University

238th AAS Meeting, June, 2021

PRESENTED AT:

1. ANALYSIS METHOD

Automatic annotation



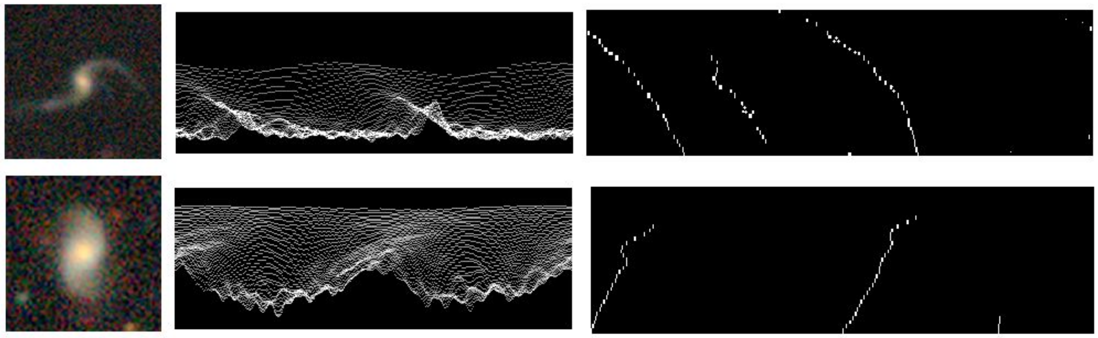
Ganalyzer turns the galaxy image into its radial intensity plot, and then applies a linear regression on the peaks identified in it. The sign of the linear regression reflects the curve of the arm, and therefore the spin direction of the galaxy.

Automatic annotation of the galaxies is required not merely due to the large size of the datasets, but also to avoid human bias. The algorithm used for the annotation is Ganalyzer (Shamir, 2011).

Ganalyzer works by clear, defined, and symmetric rules that can identify the spin direction of a spiral galaxy, but can also identify when a galaxy does not have clear spin patterns. That is different from some supervised machine learning algorithms that tend to make a forced choice to assign a certain image into a predicted class.

Ganalyzer (Shamir, 2011, 2012, 2017, 2020a,b, 2021) works by first converting a galaxy image into its radial intensity plot. The radial intensity plot is a transfomation of the galaxy image such that each pixel (x,y) in the radial intensity plot is the median of the 5x5 pixels around $(O_x + \sin(\theta) \cdot r, O_y - \cos(\theta) \cdot r)$

where θ is the polar angle and r is the radial distance from the center of the galaxy. Then the peaks in each line of the radial intensity plot are detected, and neghboring peaks are grouped to make lines. The sign of the linear regression of these lines indicate the curve of the arms, and therefore the direction towards which the galaxy rotates.



Galaxy images (left), radial intensity plots (center), and peaks in the radial intensity plots (right).

One of the most important features of Ganalyzer is that it is not based on machine learning or complex data-driven rules. It is clear how Ganalyzer works, and it does not rely on a training set or manually annotated ground truth that can capture human biases or other biases.

A detailed description of the method is available in (Shamir, 2011, 2012, 2017, 2020a,b, 2021).

Manual annotation

In addition to the automatic annotation, some of the datasets were annotated manually. To avoid bias driven by the human perception, random half of the galaxies were mirrored.

Analysis of dipole axis in the distribution of spin directions

The analysis of the distribution of the galaxies was done using X^2 statistics by fitting the cosine of the angle between the galaxy and the axis to the absolute value of the cosine of the angle multipled by the spin direction, such that a clockwise spin direction is "1" and counterclockwise spin direction is "-1" (Shamir, 2012, 2020, 2021).

That is done from each possible combination of integer RA and Dec. The X^2 computed with the actual spin directions of the galaxies is compared to the mean and standard deviation of the 1000 runs of X^2 computed when assigning the galaxies with random spin directions.

3. DESI LEGACY SURVEY

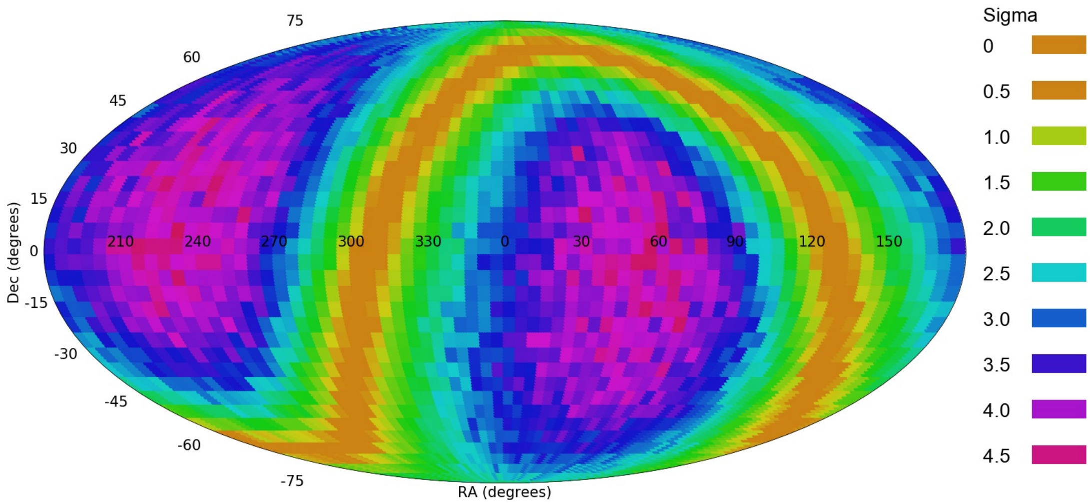
Using DESI Legacy Survey has two main advantages:

a) It is deeper and provides far more galaxies than SDSS and Pan-STARRS.

b) It covers also substantial parts of the Southern hemisphere.

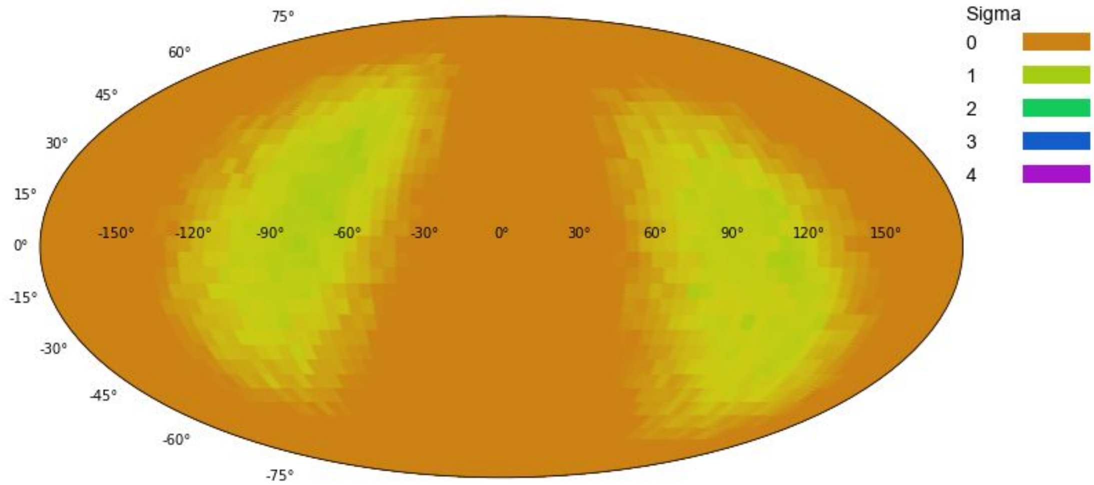
A simple analysis shows that the DESI Legacy Survey footprint can be separated into two hemispheres such that one hemisphere has a higher number of clockwise galaxies, and the other hemisphere has a higher number of counterclockwise galaxies. The bias is exactly inverse between the opposite hemispheres: 0.004 (P=0.0015) in RA (0^0-150^0 V 330^0-360^0) and -0.004 (P=0.012) in the opposite hemisphere (150^0-330^o). The analysis is based on nearly 10^6 galaxies.

Fitting the spin directions of the galaxies into cosine dependence provides a profile similar to SDSS and Pan-STARRS, and the most likely axis peaks within one sigma from the other sky surveys.



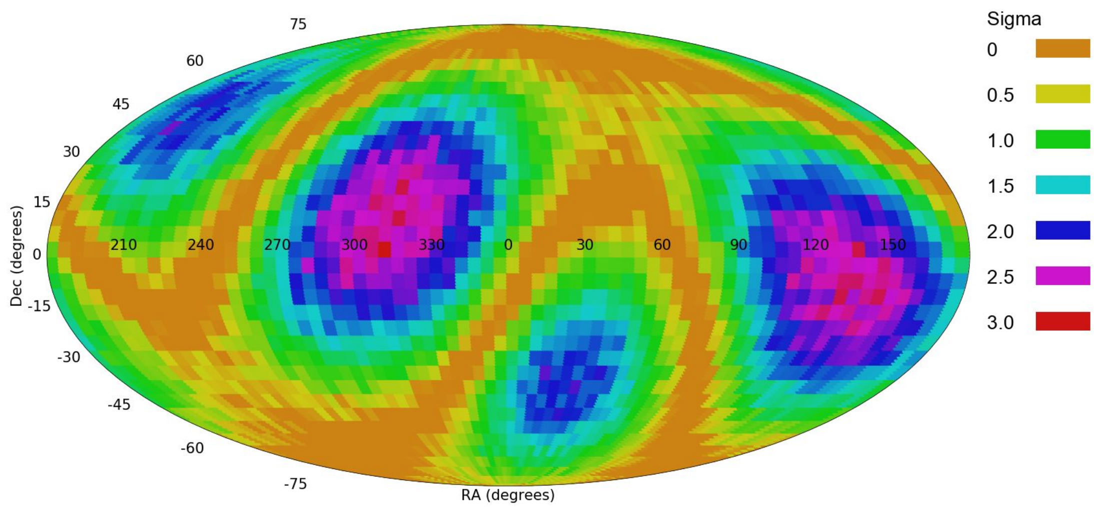
Statistical signal when fitting DESI Legacy Survey galaxies into dipole alignment.

When assigning the galaxies with random spin directions the signal is gone, and no statistically significant dipole axis can be identified.

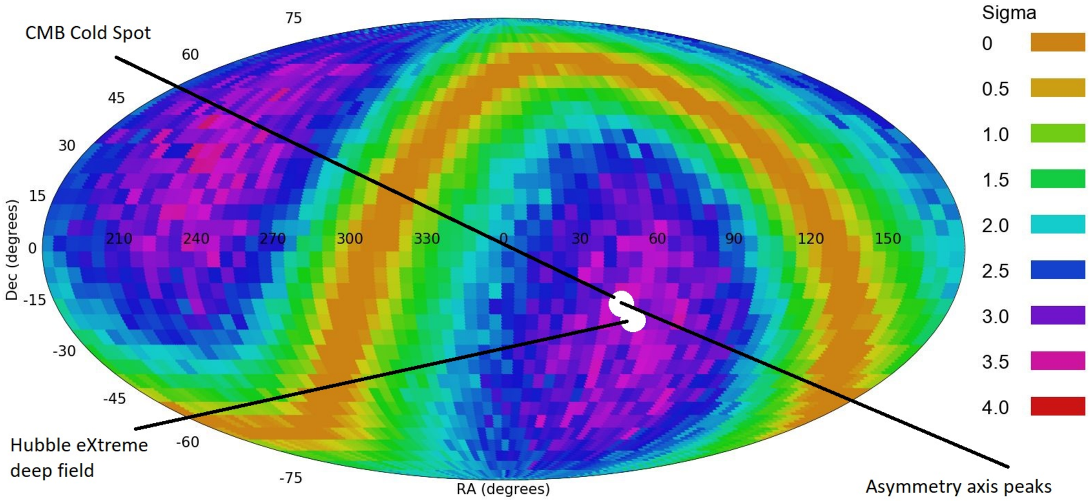


Dipole axis alignment in DESI Legacy Survey when the galaxies are assigned with random spin directions.

Fitting DESI Legacy Survey galaxies into quadrupole alignment shows somewhat weaker signal.

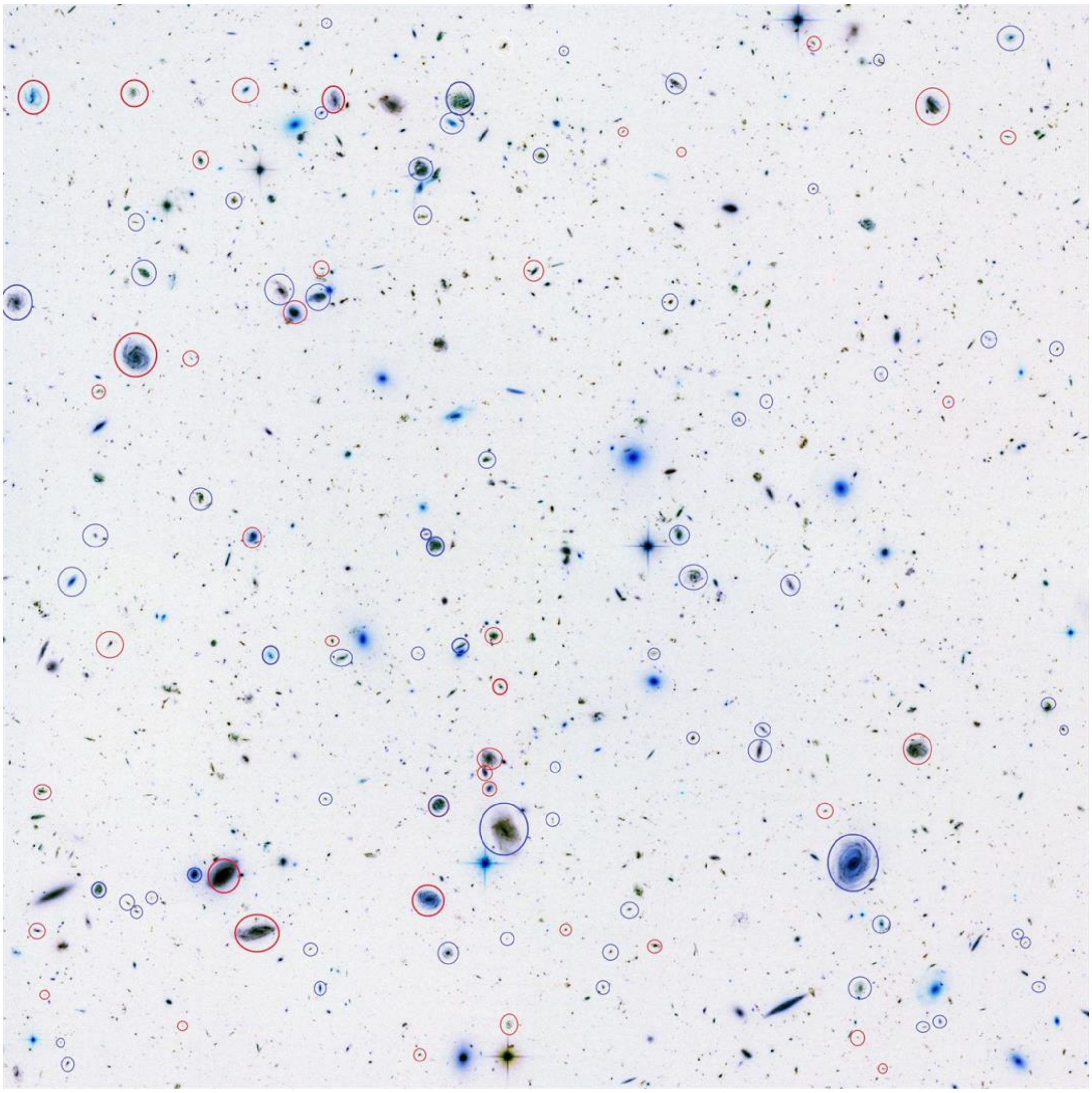


5. HUBBLE DEEP FIELDS



The most likely dipole axis, the Hubble eXtreme Deep Field, and the CMB Cold Spot (in equatorial coordinates).

Analysis of galaxies with redshift showed that the asymmetry between galaxies with opposite spin directions increases with the redshift. Based on that observation, it would be expected that a deep field at around the peak of the axis would have more galaxies spinning clockwise than counterclockwise. The closest Hubble deep field to the peak is the Hubble eXtreme Deep Field.



Hubble eXtreme Deep Field with galaxies manually annotated as clockwise (blue) or counterclockwise (red).

Counting the galaxies in the Hubble eXtreme Deep Field showed 68 galaxies spinning clockwise and 40 galaxies spinning counterclockwise. That distribution is statisticallyn significant (P=~0.0045).

This analysis has substantial weaknesses and cannot be considred a solid scientific methodology. The reasons are that the size of the data is small, and the annotation was done manually and is therefore subjective and relies on the human perception (or human imagination). The analysis with potential systematic bias is weak in any case, cartainly when the size of the data is small. Given the small size of the data and the weak analysis, this evidence cannot prove or disprove anything shown here. It agrees with the other observations, but should be considered weak evidence.

6. WHY IT IS NOT AN ERROR

The immediate explanation for an observation of this kind is an error in the analysis. Extensive work has been done to ensure not merely that no error leads to the observation, but that no error can lead to the observation. Detailed analysis of possible errors can be found in (Shamir, 2021).

1. Error in the galaxy annotation algorithm

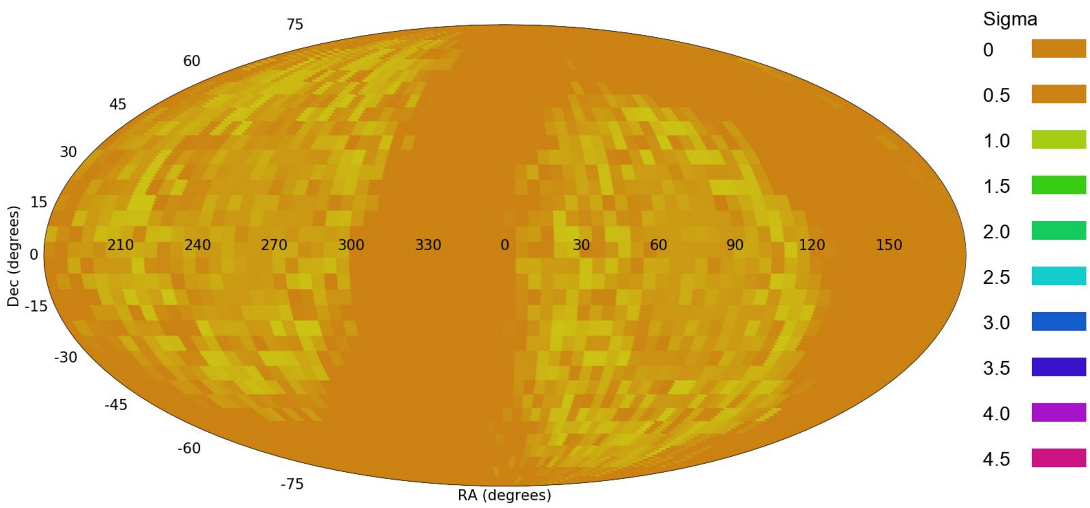
There are multiple indications that the galaxy annotation algorithm cannot lead to such observation.

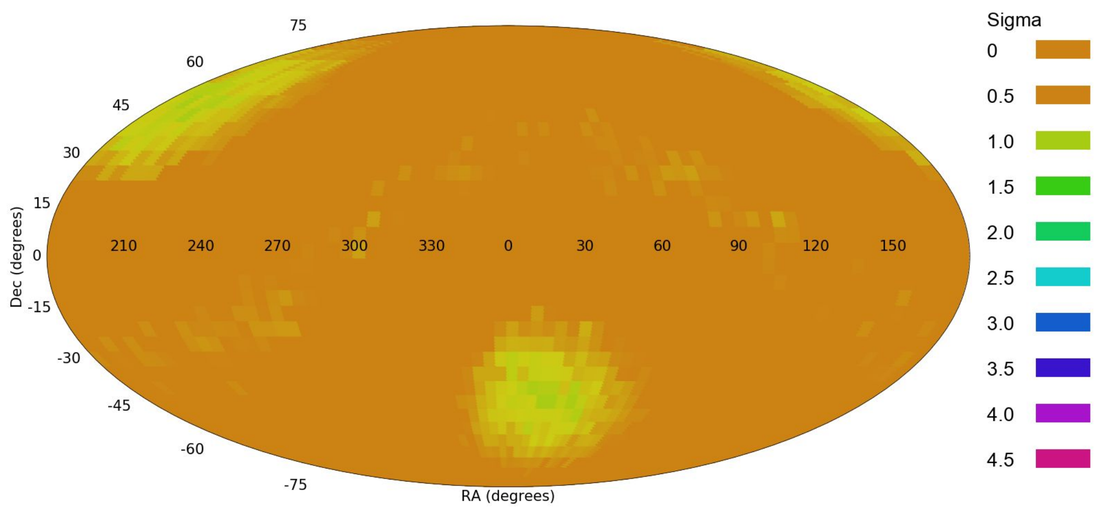
The algorithm is completely symmetric, and it is based on defined intuitive rules. It is not based on machine learning or other forms of data-driven rules that tend to be complex and non-intuitive, making it difficult to verify a system that relies on them.

Mirroring of the galaxy images provides exactly inverse results to the results when using the original images.

In any case, because each galaxy is analyzed independently, an error in the annotation would have been expected to exhibit itself in the form of consistent assymetry in all parts of the sky. Such error is not expected to "flip" in opposite hemispheres. All galaxies were analyzed by the same computer system to eliminate the possibility that the algorithm runs dofferently on different machines.

Assigning the galaxies with random spin directions immediately eliminates the signal, as shown in the figures below. The two figures attempt to fit the same galaxies of the combined dataset, but with random spin directions., to dipole and quadrupole alignment. As expected, the signal immediately disappears.





Dipole and quadrupole analysis when the galaxies are assigned with random spin directions.

Even if the algorithm is symmetric, it can still make wrong annotations.

If the galaxy annotation algorithm has a certain error the asymmetry in a certain part of the sky is defined by

$$\frac{(N_{cw}+E_{cw})-(N_{ccw}+E_{ccw})}{N_{cw}+E_{cw}+N_{ccw}+E_{ccw}}$$

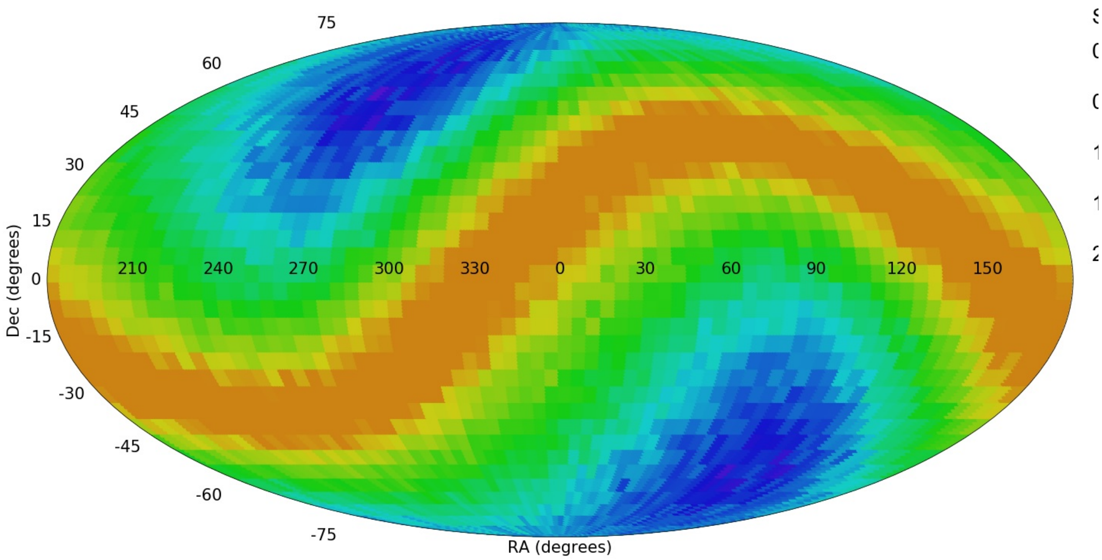
where Ecw is the number of counterclockwise galaxies annotated incorrectly as clockwise, and Eccw is the number of clockwise galaxies classified incorrectly as counterclockwise.

The galaxy classification algorithm is symmetric, and therefore the number of counterclockwise galaxies misclassified as clockwise is roughly equal to the number of clockwise galaxies missclassified as counterclockwise. Assuming Ecw=Eccw, the asymmetry can be defined as

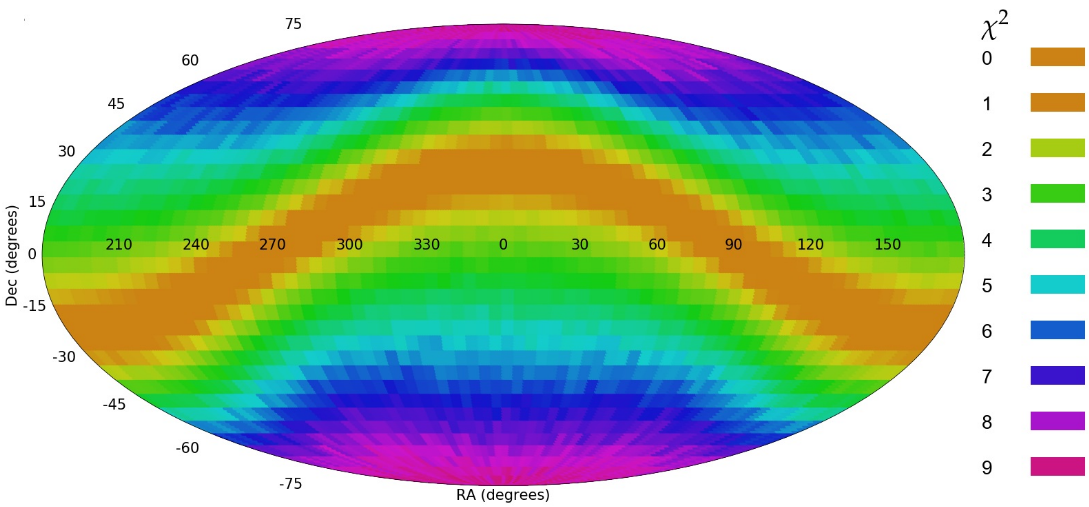
$$\frac{N_{cw}-N_{ccw}}{N_{cw}+E_{cw}+N_{ccw}+E_{ccw}}$$

Because Ecw and Eccw cannot be negative, misclassified galaxies are expected to make the asymmetry lower. Therefore, misclassified galaxies are not expected to lead to the observed asymmetry.

Several experiments were used to test the impact of misclassified galaxies empirically. The figure below shows SDSS data such that 25% of the galaxies were assigned with random spin directions (Shamir, 2021).



The figure shows that the error did not have substantial impact on the results, mainly because it is symmetric (Shamir, 2021). However, when the error is not symmetric, even a small 1% error leads to very strong bias, and a dipole axis that peaks exactly at the celestial pole



Additionally, the analysis of some of the datasets was based on manual annotation of the galaxies, with no use of automation.

2. Bias in the sky survey hardware or photometric pipeline

Autonomous digital sky surveys are complex research instruments. It is difficult to think of an error in the hardware or software that can lead to asymmetry between the number of clockwise and counterclockwise galaxies, but due to the complexity of these systems it is also difficult to prove that such error does not exist.

However, while it is difficult to think of such error in one telescope, it is clearly difficult to think of such error that happen consistently in four telescopes.

3. Cosmic variance

The asymmetry is determined by the difference between two measurements made in the same field therefore the probe of asymmetry between the number of clockwise and counterclockwise galaxies observed from Earth is a relative measurement. Therefore, that measurement is not be affected by cosmic variance. Any effect that impacts the number of clockwise galaxies observed from Earth is expected to have a similar effect on the number of counterclockwise galaxies.

4. Multiple photometric objects in the same galaxy

In some cases, digital sky survey can identify several photometric objects as independent galaxies, even in case they are part of one large extended object. In the datasets used here all photometric objects that are part of the same galaxy were removed by removing all objects that had another object within 0.01^{o} away.

Even of such objects existed in the dataset, they are expected to be evenly distributed between galaxies that spin clockwise and galaxies that spin counterclockwise. Experiments by using datasets of galaxies assigned with random spin directions and adding artificial objects to the galaxies showed that adding objects at exactly the same position of the original galaxies does not lead to signal of asymmetry (Shamir, 2021).

5. Atmospheric effect

There is no know atmospheric effect that can make a galaxy that spin clockwise appear as if it spin counterclockwise. Also, because the asymmetry is always measured with galaxies imaged in the same field, any kind of atmospheric effect that affects galaxies the spin clockwise will also affect galaxies that spin counterclockwise. Therefore, it is unlikely that a certain atmospheric effect would impact the number of clockwise galaxies at a certain field, but would have different impact on galaxies spinning counterclockwise. In any case, one of the datasets used here is made of galaxies imaged by the space-based Hubble Space Telescope, and are therefore not subjected to any kind of atmospheric effect.

6. Backward spiral galaxies

Backward spiral galaxies are relatively rare, and expected to be distributed equally between galaxies with different spin directions. Therefore, there is no reason to assume that the observations shown here are driven by backward spiral galaxies.

2. COMPARISON OF 11 DATASETS

The data was acquired from several different telescopes, including SDSS, Pan-STARRS, HST, and DESI Legcay Survey. Some of the datasets were classified manually, and some were classified automatically. The table summarizes the distribution of the galaxy spin directions.

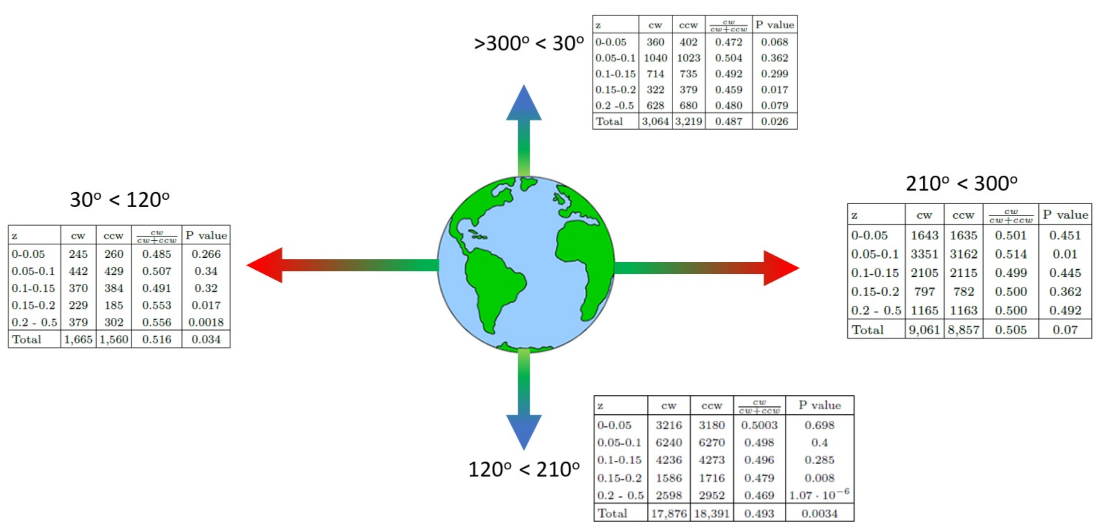
	Dataset	Telescope	# galaxies	z	RA	Dec	Annotation	sigma
1	Shamir, 2020b	SDSS	38998	0.04	229	-21	Automatic	<2
2	Shamir, 2020b	Pan-STARRS	33028	0.04	227	1	Automatic	<2
3	Longo, 2011	SDSS	15158	0.05	217	32	Manual	5.15
4	Shamir, 2016	SDSS	13440	0.06	165	30	Manual	4.02
5	Shamir, 2021	SDSS	77840	0.07	165	40	Automatic	2.56
6	Galaxy Zoo	SDSS		0.07	161	15	Manual	<2
7	Shamir, 2012	SDSS	126501	0.08	135	32	Automatic	4.27
8	Shamir, 2017	SDSS	40739	0.09	120	27	Manual	2.4
9	Shamir, 2020b	SDSS	63693	0.12	69	56	Automatic	4.63
10	Shamir, 2020c	SDSS	15863	0.25	71	61	Automatic	7.38
11	Shamir, 2020c	HST	8690	0.58	78	47	Manual	2.8

Most likely dipole axis identified in 11 different datasets, including three different telescopes and both automatic and manual annotations of the galaxies.

The analysis shows that datasets with similar mean redshift of the galaxies also show dipole axes peaking at around the same part of the sky. With the exception of datasets 1, 2 and 6, all datasets show statistically significant dipole axis.

Redshift dependence

The following figure shows a simple analysis of counting SDSS galaxies spinning in opposite directions at different RA and redshift ranges (Shamir, 2020b). It is done with ~6.4*10^4 SDSS galaxies with redshift.



>300° < 30°

z	cw	ccw	$\frac{cw}{cw+ccw}$	P value
0-0.05	360	402	0.472	0.068
0.05-0.1	1040	1023	0.504	0.362
0.1-0.15	714	735	0.492	0.299
0.15-0.2	322	379	0.459	0.017
0.2 -0.5	628	680	0.480	0.079
Total	3,064	3,219	0.487	0.026

30° < 120°

z	cw	ccw	$\frac{cw}{cw+ccw}$	P value
0-0.05	245	260	0.485	0.266
0.05-0.1	442	429	0.507	0.34
0.1-0.15	370	384	0.491	0.32
0.15-0.2	229	185	0.553	0.017
0.2 - 0.5	379	302	0.556	0.0018
Total	1,665	1,560	0.516	0.034

210° < 300°

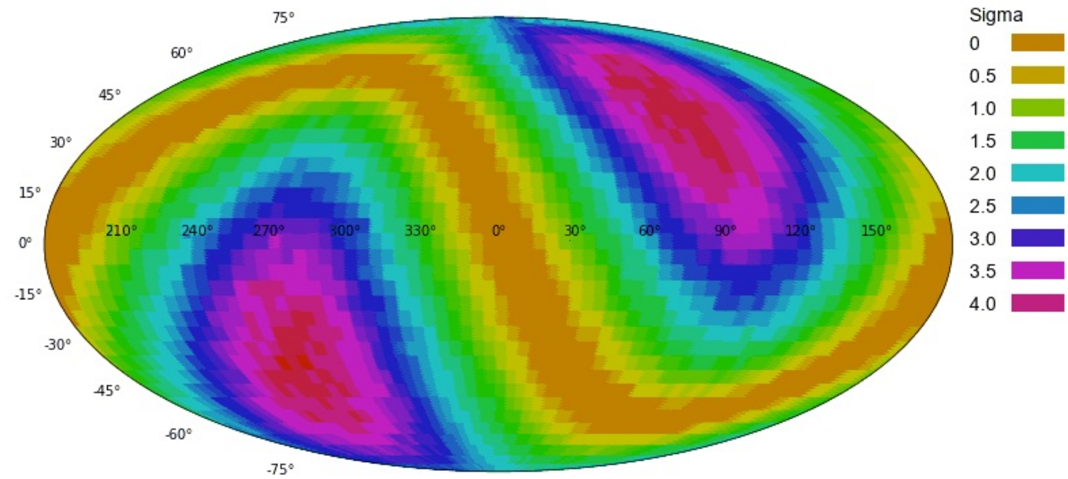
z	cw	ccw	$\frac{cw}{cw+ccw}$	P value
0-0.05	1643	1635	0.501	0.451
0.05-0.1	3351	3162	0.514	0.01
0.1-0.15	2105	2115	0.499	0.445
0.15-0.2	797	782	0.500	0.362
0.2 - 0.5	1165	1163	0.500	0.492
Total	9,061	8,857	0.505	0.07

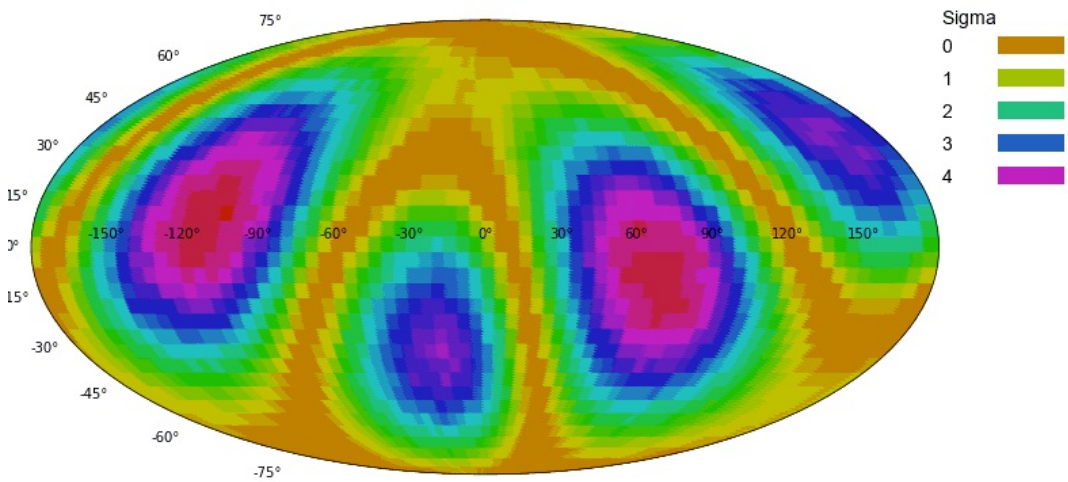
120° < 210°

z	cw	ccw	$\frac{cw}{cw+ccw}$	P value
0-0.05	3216	3180	0.5003	0.698
0.05-0.1	6240	6270	0.498	0.4
0.1-0.15	4236	4273	0.496	0.285
0.15-0.2	1586	1716	0.479	0.008
0.2 - 0.5	2598	2952	0.469	$1.07 \cdot 10^{-6}$
Total	17,876	18,391	0.493	0.0034

Distribution of the spin directions in different RA and redshift ranges. The analysis was done with 6.4*10^4 SDSS galaxies with redshift (Shamir, 2020b).

Fitting each point in the sky with dipole and quadrupole alignment provide the following profiles.



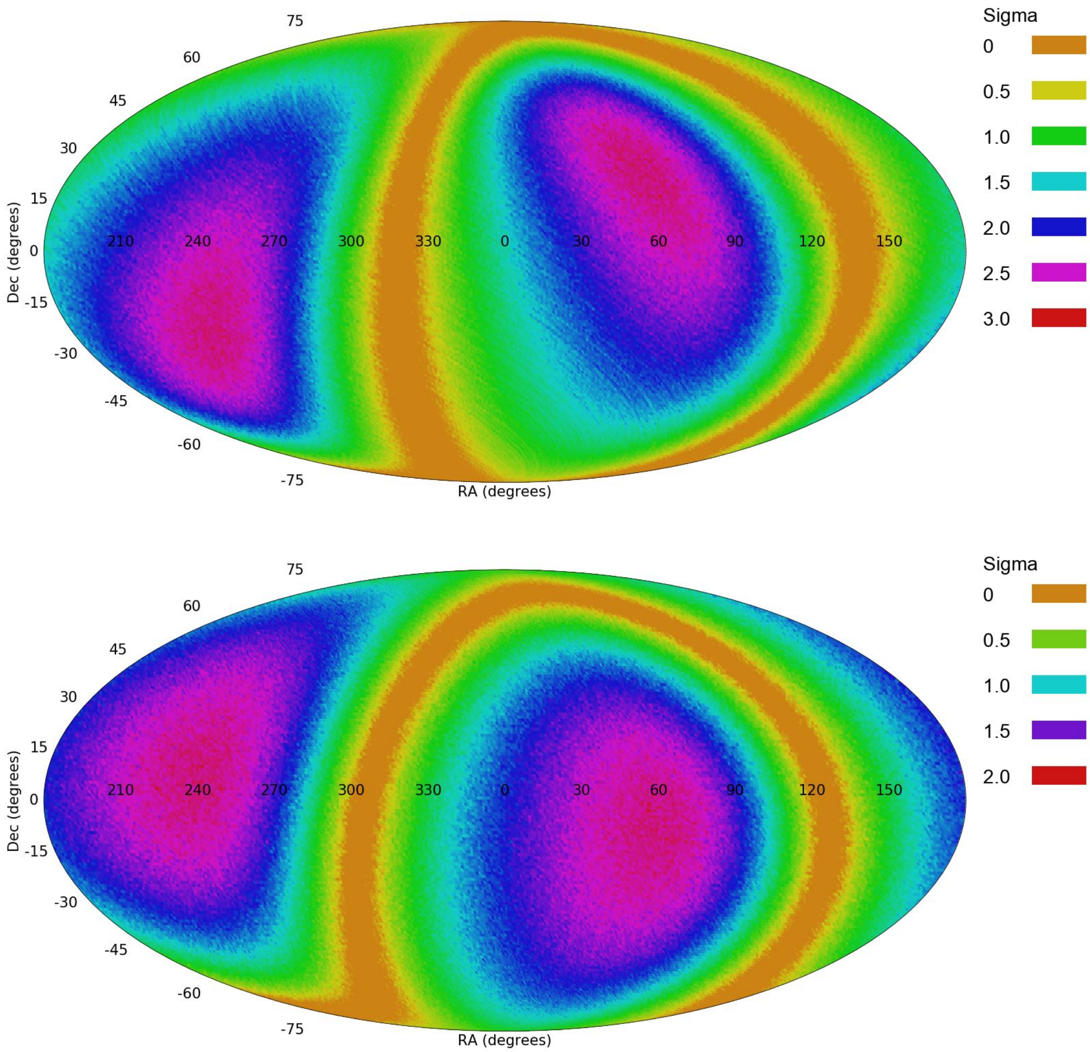


Dipole and quadrupole alignment from each point in the sky using SDSS galaxies (Shamir, 2020b).

Comparison of different telescopes

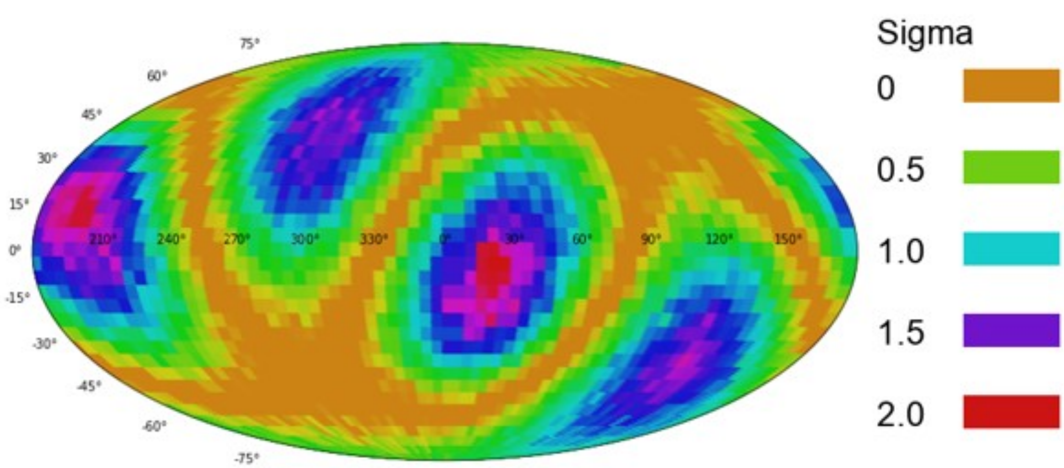
When normalizing for the redshift, the analysis using different telescopes shows very similar results (Shamir, 2020a,b).

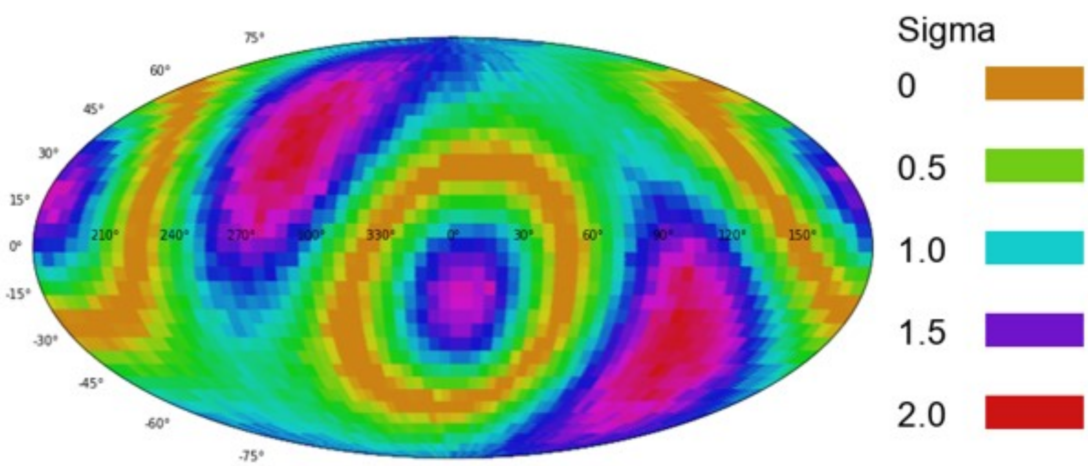
For instance, the profile of dipole alignment and quadrupole alingment in SDSS and Pan-STARRS is very similar when the two datasets have similar redshift distribution (Shamir, 2020b). Pan-STARRS galaxies do not have redshift, so their redshift distribution was profiled by using a sample of galaxies that have redshift in SDSS.



SDSS (top) and Pan-STARRS proifle of dipole alignment.

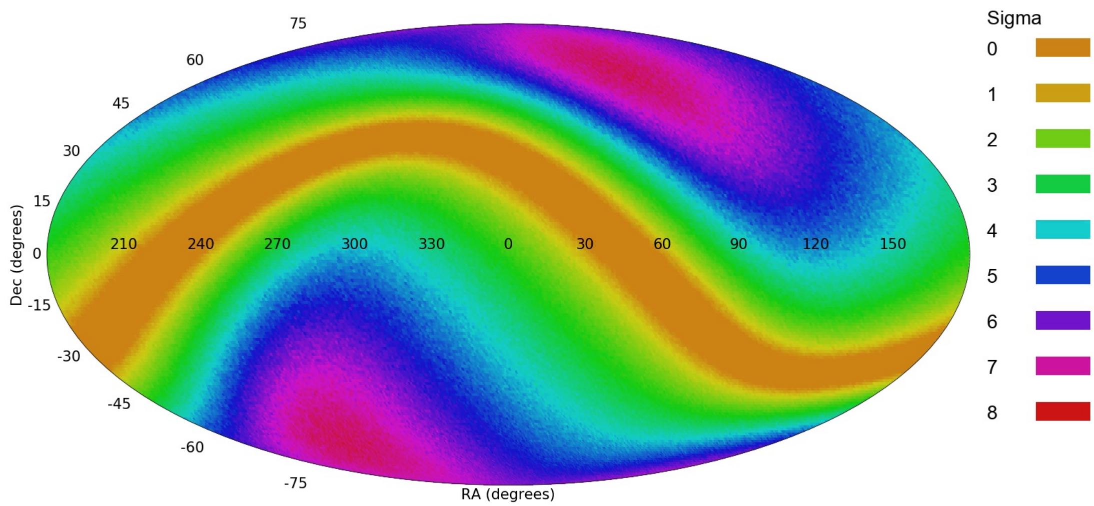
Same agreement is also observed with quadrupole alignment.

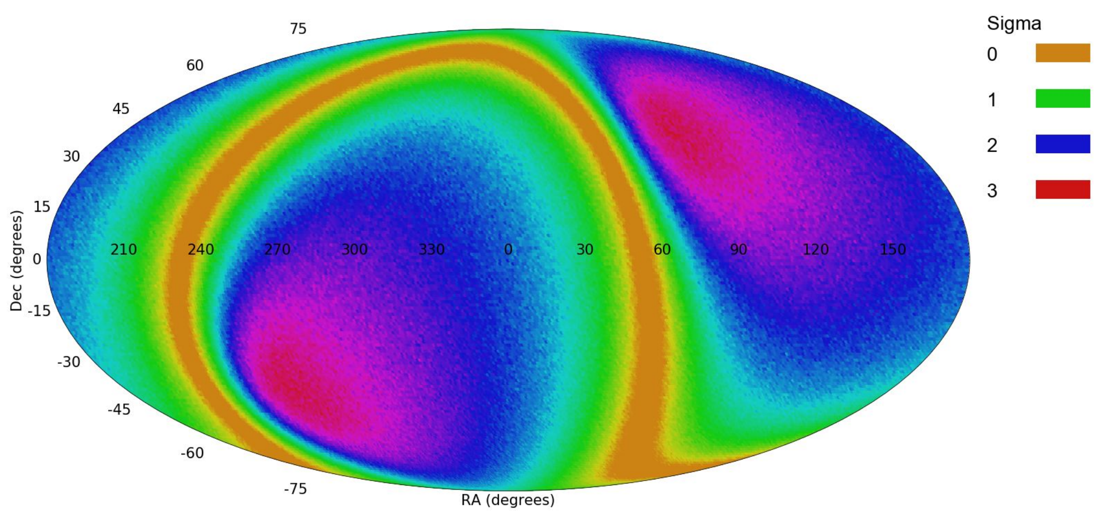




Quadrupole alignment in SDSS (top) and Pan-STARRS galaxies when the redshift distribution in both datasets is similar.

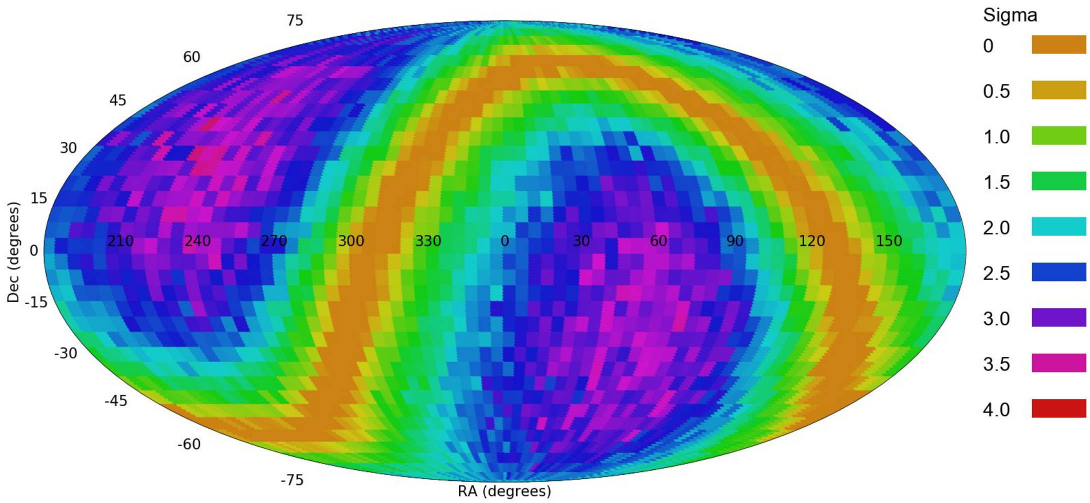
The profile of asymmetry observed with 8,690 HST galaxies is similar to the profile observed with SDSS data such that only galaxies with z>0.15 are used.



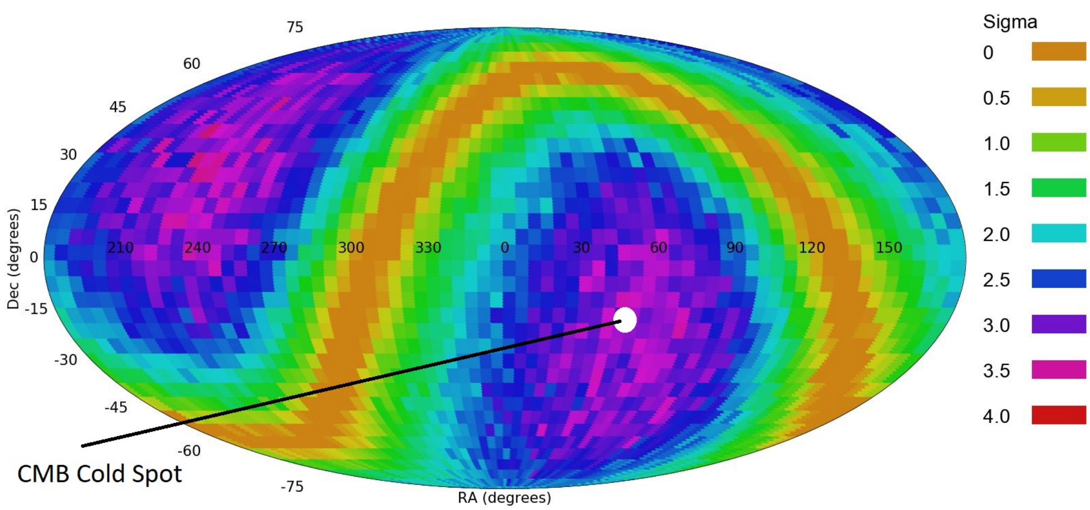


Dipole alignment of the spin directions of 8,690 HST galaxies (bottom) and SDSS galaxies with z>0.15.

4. "META ANALYSIS" - COMBINING NORTHERN AND SOUTHERN HEMISPHERES



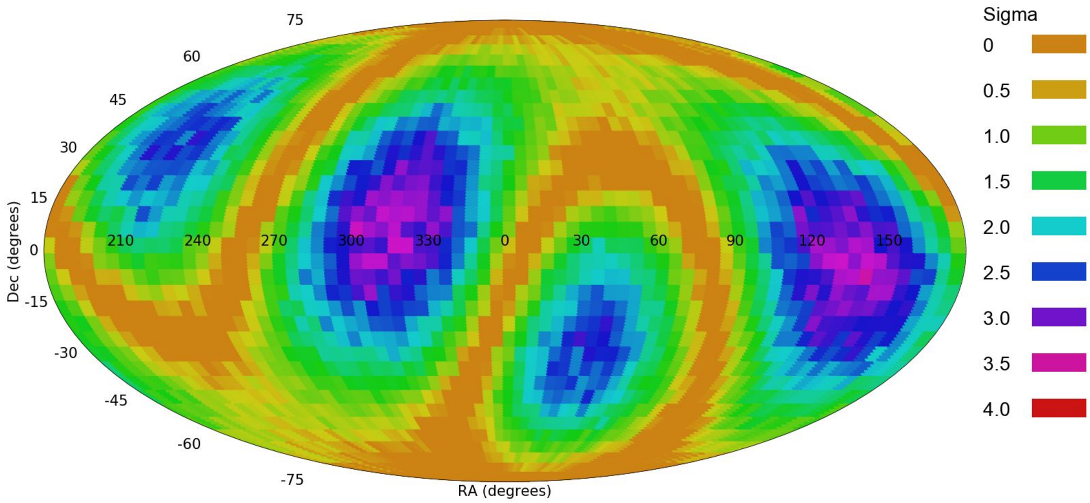
Combining all data from SDSS, HST, Pan-STARRS, and DESI Legacy Survey provides a large dataset that covers both the Northern and Southern hemispheres. The provides a similar profile observed with the SDSS and Pan-STARRS galaxies, showing that different telescopes provide the same results, regardless of whether the data was acquired in the Northern or Southern hemispheres. That allows more accurate analysis of the most likely dipole axis. Mainly, it allows to better estimate the location of the declination, as the declination range is limited when using just one hemisphere. The most likely dipole axis is identified at (RA=47,Dec=-22). Interestingly, that is very close to the CMB Cold Spot.



The location of the CMB Cold Spot compared to the location of the most likely dipole axis when using data from all four sky surveys, and from both the Northern and Southern hemispheres. Note that the projection above uses equatorial coordinates.

Question: Why is the axis in the distribution of galaxy spin directions aligned with the CMB Cold Spot? Is it a coincidence? Could there be a reason for that?

Fitting the galaxy spin directions to quadrupole alignment from all possible (RA, Dec) integer combinations provides the following profile. Also with quandrupole alignment, one of the axes is with close proximity to the CMB Cold Spot. The statistical significance of the quadrupole alignment is somewhat weaker than the dipole alignment.



Fitting the galaxy spin directions of the four sky surveys to quadrupole alignment from all possible (RA, Dec) combinations.

7. IDEAS THAT SHOULD NOT BE DONE

The analysis of the distribution of galaxy spin directions should be done carefully, and with sound experimental design. Below are several real-world bad ideas that can lead to different or unpredictable results.

Bad idea 1: 3D analysis using the photometric redshift

The redshift computed from the photometry data is NOT the same as the redshift measured from the spectra. The photometric redshift is inaccurate, and can be systematically biased. The best photometric redshift algorithms have an error of 10%-20%. Using data with 10%-20% error to identify 1%-2% bias is unintelligent, and essentially leads to a random number generator. The photometric redshift cannot be used for this purpose. Any attempt for 3D analysis with the photometric redshift is meaningless.

Bad idea 2: Limiting to low redshift ranges

Analysis with the redshift (not the photometric redshift) shows that the signal of the asymmetry increases as the redshift gets hgiher (Shamir, 2020b). It's not necessarily a bad idea to examine low redshifts, but it should be done with the understanding that there is much weaker signal at ($z<0.15$).

Bad idea 3: Using Galaxy Zoo annotation

Galaxy Zoo is a fantastic successful venture that connects the broader community with science. While Galaxy Zoo can be credited for many successful achievements, the ability to use human volunteers to annotate the spin directions of galaxies is not one of them. Evidently, the human brain is not a reliable system when it comes to annotation of the spin patterns of spiral galaxies.

The initial attempt made by Galaxy Zoo to use crowdsourcing to annotate spin directions of galaxies was reasonable. The biases that they found were not known before Galaxy Zoo started. But they are known now. Galaxy Zoo annotations of clockwise and anti-clockwise spins are inaccurate, and includes very substantial systematic biases. The most concerning aspect of using Galaxy Zoo data for this purpose is not that the annotations are inaccurate, but that the bias is systematic. Galaxy Zoo annotations therefore cannot be considered "ground truth". Galaxy Zoo annotations cannot be used for identifying subtle biases, and can lead to unpredicted results that may or may not have astronomical meaning.

Bad idea 4: Using machine learning/pattern recognition/deep learning to annotate galaxies

Machine learning, and specifically deep learning, provide effective solutions to the analysis of very large databases of galaxy images. However, while they are effective and easy to use, they are not suitable for a study of this kind. These algorithms are trained to discriminate between the classes by any piece of information they can learn from the training data. That leads to complex data-driven unintuitive rules that are virtually impossible to verify their correctness and ensure absence of any kind of bias. For instance, they also learn from the background, and can provide different results if they are trained with a different set of galaxies. Even the order of the training samples changes the rules and the performance of some machine learning algorithms such as deep neural networks. These changes are subtle, but these subtle changes are sufficient to lead to completely different results.

For these reasons, convolutional neural networks are virtually impossible to be fully symmetric. Even if the training data is exactly the same, the order by which the network is being trained can change the way the network behaves. The initial weights add another potential level of randomness. Because the network also learns the image background, the asymmetry can changed based on the part of the sky. Convolutional neural networks are not designed and cannot be used for this purpose.

Supervised machine learning algorithms also make a forced-choice, and associate each sample with a certain class. Many galaxies do not have identifiable spin direction, and their distribution can greatly affect the results. Obviously, it is possible to make a class of "unidentifiable" spin directions, but that brings back the problem of the symmetricity of the algorithm, which is virtually impossible to verify when using machine learning.

AUTHOR INFO

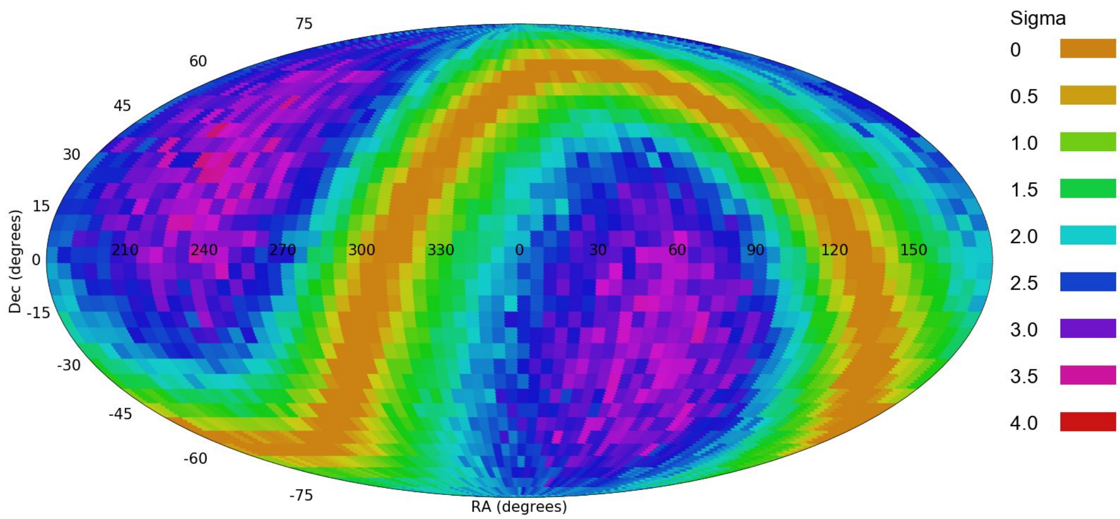
Lior Shamir

Kansas State University

http://people.cs.ksu.edu/~lshamir/

ABSTRACT

Several previous studies have suggested that the distribution of galaxy spin directions is not fully random, and exhibits patterns of asymmetry at cosmological scales. Here I compare 12 different datasets of spiral galaxies sorted by their spin directions. Analysis of the large-scale distribution of galaxy spin directions in each of the datasets shows large-scale dipole axes at different levels of statistical signal. While the direction of the most likely axis is somewhat different in different datasets, datasets with similar redshift distribution of the galaxies show similar directions of the dipole axes. Moreover, the change in the direction of the axis correlates with the changes in the mean redshift of the galaxies in each dataset. The datasets were collected by four different instruments: *SDSS*, *Pan-STARRS*, *HST*, and *DESI Legacy Survey*. The galaxies in each dataset were classified by their spin directions either manually or automatically. The different datasets show similar distribution and similar location of the most likely dipole axis in similar redshift ranges. The datasets also show a correlation between the direction of the dipole axis and the mean redshift of the galaxies. Combining data from both Northern and Southern hemispheres provides a more comprehensive analysis compared to using just one hemisphere. Interestingly, the most likely dipole axis peaks at very close proximity to the CMB Cold Spot.



REFERENCES

Dojcsak, L., Shamir, L., Quantitative analysis of spirality in elliptical galaxies, *New Astronomy*, 28, 1-8, 2014.

Longo, M., Physics Letters B, 699, 224, 2011.

Shamir, L., Analysis of the alignment of non-random patterns of spin directions in populations of spiral galaxies, *Particles*, 4(1), 11-28, 2021a.

Shamir, L., Galaxy spin direction distribution in HST and SDSS show similar large-scale asymmetry, *Publications of the Astronomical Society of Australia*, 37, e053, 2020a.

Shamir, L., Patterns of galaxy spin directions in SDSS and Pan-STARRS show parity violation and multipoles, *Astrophysics and Space Science*, 365, 136, 2020b.

Shamir, L., Large-scale asymmetry between clockwise and counterclockwise galaxies revisited, *Astronomical Notes*, 341(3), 324-330, 2020c.

Shamir, L., Asymmetry between galaxies with different spin patterns: A comparison between COSMOS, SDSS, and Pan-STARRS, *Open Astronomy*, 29, 2020e.

Shamir, L., Large-scale photometric asymmetry in galaxy spin patterns, *Publications of the Astronomical Society of Australia*, 34, e044, 2017a.

Shamir, L., Colour asymmetry between galaxies with clockwise and counterclockwise handedness, *Astrophysics and Space Science*, 362, 33, 2017b.

Shamir, L., Handedness asymmetry of spiral galaxies with z<0.3 shows cosmic parity violation and a dipole axis, *Physics Letters B*, 715, 25-29, 2012.

Shamir, L., Color differences between clockwise and counterclockwise spiral galaxies, *Galaxies*, 3(1), 215-220.

Shamir, L., Ganalyzer: A tool for automatic galaxy image analysis, *Astrophysical Journal*, 736(2), 141, 2011.

Shamir, L., Asymmetry between galaxies with clockwise handedness and counterclockwise handedness, *Astrophysical Journal*, 823(1), 32, 2016.