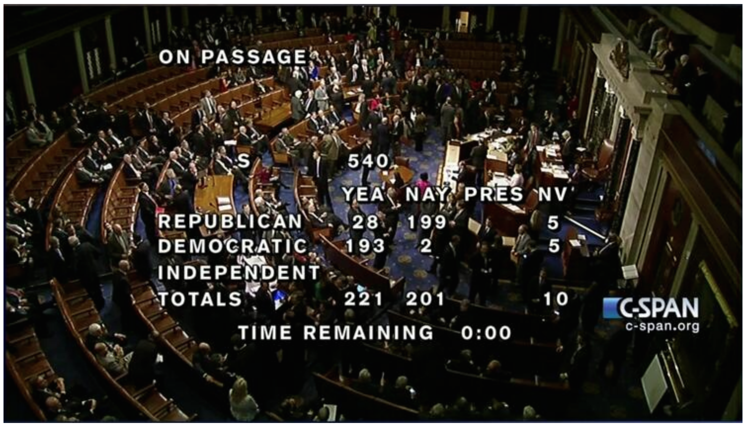
Supporting Information (to go online) for: Using Motion Detection to Measure Social Polarization in the U.S. House of Representatives.

Bryce J. Dietrich*

Contents

*Department of Political Science, University of Iowa, 341 Schaeffer Hall, Iowa City, IA 52242 (bryce-dietrich@uiowa.edu, http://www.brycejdietrich.com).

Figure S1: Overhead Shot of Members of Congress Mingling after a Roll-Call Vote on the House Floor



Note: "Ant farm" shots similar to this appear frequently on C-SPAN. Not only do they show all of the social interactions that take place after a floor vote, but they have been a quintessential part of C-SPAN coverage. All the analyses presented below considers videos similar to the frame shown here.

S1 Data and Methods

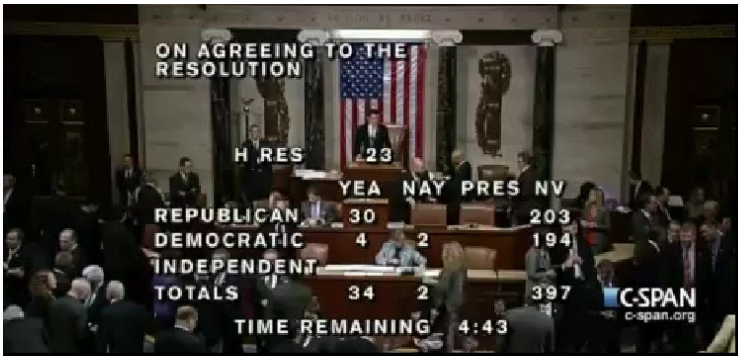
S1.1 C-SPAN Video Data

I collected 6,526 C-SPAN videos for this study. Each video was approximately 16 minutes long with the first video occurring on January 7, 1997 and the last video occurring on December 13, 2012. Although these videos can be used to measure a variety of things, I focused my efforts on understanding the social interactions that take place in shots similar to the overhead shot shown in Figure S1 which I call the "ant farm" shot.

First, this shot is a quintessential example of what is found on C-SPAN. While there is considerable variation in some of the close-up shots C-SPAN has used over time, this shot has been a consistent part of their broadcasts. Indeed, anyone who watches C-SPAN on a regular basis has likely seen this shot which is why it is particularly interesting to anyone who wants to better understand what we can learn from C-SPAN coverage. Moreover, the vast majority of floor votes include this shot making it well-suited for the present paper.

Second, the present study is interested in aggregate patterns of social interactions. The "ant farm" shot shows *all* the interactions that occur after floor votes, while other camera angles only show a portion of these discussions. For example, C-SPAN often broadcasts images similar to the one shown in Figure S2. Even though it is possible to identify some members of Congress (MCs) in this shot, it only shows a small part of the well of the House. Conversely, the "ant farm" shot shows all the interactions that take place after floor votes, making it particularly useful for understanding whether Democrats and Republicans generally talk with one another.

Figure S2: Close-Up Shot of Members of Congress Mingling after a Roll-Call Vote on the House Floor



ON AGREEING TO THE RESOLUTION

H RES 23

	YEA	NAY	PRES	NV
REPUBLICAN	30			203
DEMOCRATIC	4	2		194
INDEPENDENT				
TOTALS	34	2		397

TIME REMAINING 4:43

Note: C-SPAN also frequently broadcasts the well of the House. These "close up" shots only show a fraction of the social interactions that take place, whereas the "ant farm" shots broadcasts the whole floor. In Section S1.1, I introduce a method for differentiating between these two images.

To collect the initial data, I worked with Alan Cloutier at the C-SPAN video library. He used C-SPAN's timestamps to create videos after each floor vote. The videos start with the gavel calling for the vote to begin and end with the next gavel – indicating the vote has concluded. This means the videos contain many different types of shots, ranging from close-ups of the Speaker of the House to the overhead shots used in this study. This creates a proverbial needle in a haystack where the needle is the specific shots needed for this study and the haystack is the millions of frames that appear in the raw C-SPAN corpus obtained for this study. The algorithm I developed to overcome this problem is described in Section S1.3 of the Supplemental Information (SI). In the next subsection, all of the variables used in this study are explained, which is where I now turn.

S1.2 Independent and Dependent Variables

Due to page limits, I was unable to explain many of the variables reported in Table 1 in the main text. Table S1 provides a detailed description of all variables which should help readers better understand why they were included in the model. Generally speaking, almost all of the variables are calculated at the vote-level, except for variables related to the sponsor which were calculated at the bill-level. Finally, both the Congress fixed-effects and the dummy variable indicating whether it was an election year both capture some of the temporal dynamics.

Table S1: Variable Descriptions

Variable	Description	Source	Level
Structural Similarity	A continuous variable indicating the degree to which a video had motion. The variable is described on page 6 in the main text. It is also described at great length on pages S9–S12 in the SI. The variable was created using frame differencing and the data described on pages S6 – S9 in the SI.	*C-SPAN*	Vote
Previous Party Votes	A continuous variable capturing whether partisan voting is generally occurring irrespective of the social interactions observed in the C-SPAN videos. More specifically, this is the percent of previous votes in which the majority of Democrats cast votes in the opposite direction of the majority of Republicans. More details can be found on page 9 of the main text.	*Voteview*	Vote
Passage Vote	A dichotomous variable indicating whether the vote was on a passage question. This was included because passage votes are generally more important which may influence the social dynamics immediately afterwords.	*Voteview*	Vote
Amendment Vote	A dichotomous variable indicating whether the vote was on an amendment. This was included because Fowler (2006) used amendments as one of his main dependent variables. To him, passing amendments was a signal of individual influence. Given that, I included this variable as a control since such actions could also influence the legislative environment.	*Voteview*	Vote
Not Voting	A continuous variable indicating the number of MCs who did not vote. This variable was used as another way to measure the importance of a vote. Just as passage votes are generally more important, votes in which large numbers of MCs do not vote are likely less consequential. Again, I suspect this may influence the social dynamics immediately after the vote occurs.	*Voteview*	Vote

|Sponsor Ideology|	The absolute value of the bill sponsor's DW-Nominate score. Generally speaking, ideologically extreme bills are more likely to be divisive which could influence the social interactions afterwords. Since it is difficult to measure such extremity directly, I use the absolute value of the sponsor's DW-Nominate score as a useful proxy. This continuous variable ranges from 0 (less extreme) to 1 (more extreme).	*Voteview* and *CBP*	Bill
Sponsor Seniority	A continuous variable indicating the years served by the bill's sponsor. As suggested by a colleague, more senior MCs likely have a better sense of the legislative environment. Given that, the bills they sponsor could theoretically carry more weight in terms of the social interactions that occur immediately afterwords. Moreover, due to their tenure in the legislature, bills sponsored by more senior MCs are also likely to be more important and consequently more socially influential.	*CBP*	Bill
Sponsor Party Leader	A dichotomous variable indicating whether the bill sponsor held any of the following positions: Speaker of the House, Majority/Minority Leader or Majority/Minority Whip. Similar to `Sponsor Seniority`, this variable was included since bills sponsored by party leaders are likely to be more important and consequently more socially influential.	*CBP*	Bill
Election Year	Since legislative behavior is known to change when MCs are worried about re-election, I also included a dichotomous variable which returns a 1 when the social interactions observed on C-SPAN occur in the same year as an election. This was done for each video.	*CBP*	Year
Congress	As shown in Figure 3 in the main text, bipartisan social interactions are increasingly less likely to occur as time progresses. To help control for this temporal dynamic, dichotomous variables were included for each Congress with the 105th Congress serving as the baseline category. This was done for each video.	*CBP*	Year

All of the data used in this study was derived from three primary sources: *C-SPAN*,[S1] *Voteview*,[S2] and *Congressional Bills Project* (CPB).[S3] For the most part the non-video measures are straightforward and control for a number of factors which may create a more polarized legislative environment. For more details about each variable, please consult Table S1 which describes each variable and highlights some of the more relevant sections of the main text and SI. In the following subsections, I explain each of the video measures created for this study.

S1.3 Video Fingerprinting

Figure S3 shows how the "ant farm" shots were extracted from the C-SPAN videos. A research assistant first manually identified 17,700 "good" frames using a random sample of videos. I then used a video hashing (or fingerprinting) algorithm developed Zauner (2010) to compare each frame to every other frame. A high-performance computing cluster then calculated 113,935,802,380 pairwise comparisons. Frames that shared at least 10 of 16 hexadecimal characters (after hashed) with at least one of the "good" frames were said to include the overhead shot. Each of these steps are explained in more detail below.

Step 1: Extracting greyscale frames. I first used **ffmpeg** to create a single frame for each second of the 6,526 C-SPAN videos, ultimately yielding 6,411,694 frames. Since each image represents a matrix of pixels equivalent to the frame dimensions, it is often easier to work in grayscale. For example, the frame found in Figure S1 is composed of three matrices which correspond to the distribution of red, green, and blue with a 0 representing the total absence of that color and a 1 representing the total presence. To convert the image to grayscale, I simply took the average across all three channels which yields a 0 when there is no color (black) and a 1 when all three colors are present – also known as white.

Step 2: Finding seed frames. In order to identify the "good" frames, I took a 1 percent sample of all 6,411,694 grayscale frames. Once obtained, I reviewed all the frames and selected 5 seed frames for each Congress which looked essentially identical to Figure S1, but for some minor variations. For example, some of the frames included a text overlay of the vote counts, whereas others did not. I then used these frames to generate a larger list of potential "good" frames that were ultimately used to score the rest of the frames.

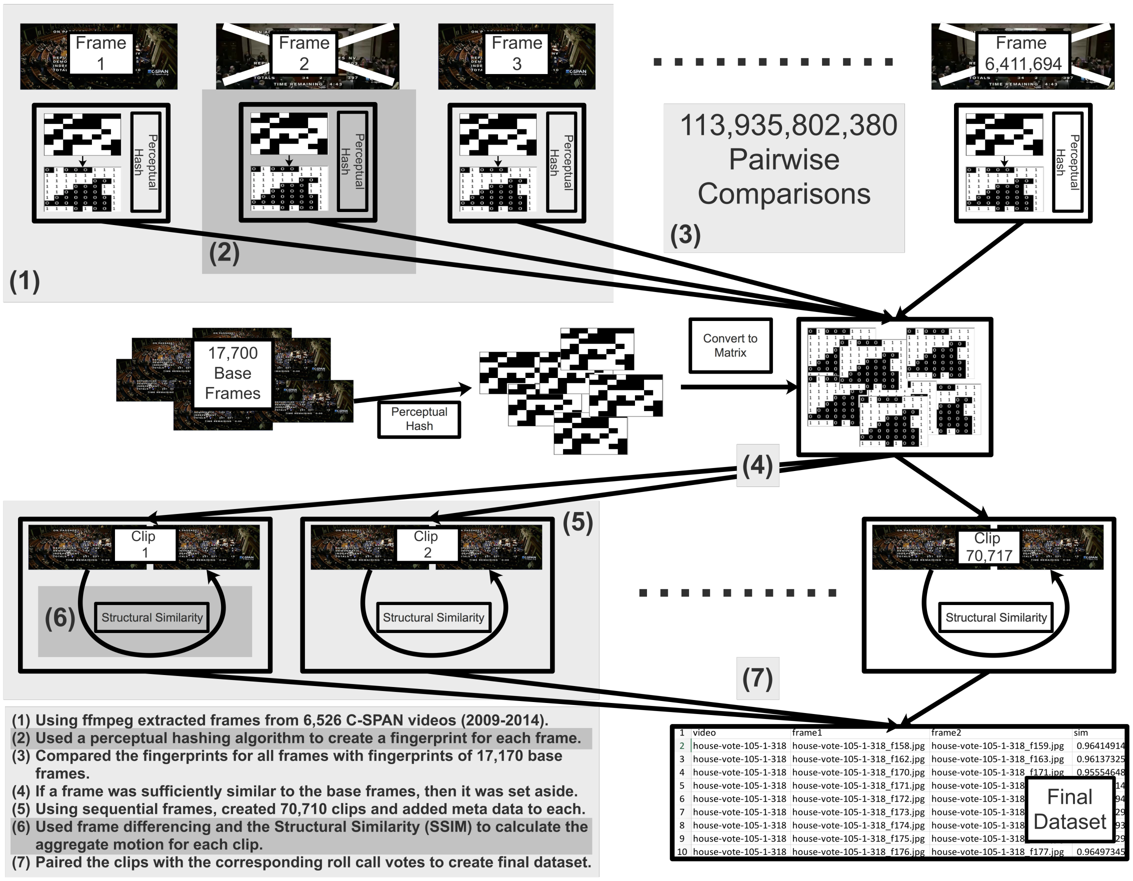
Step 3: Fingerprinting each frame. The "good" frames were identified by comparing all 6,411,694 grayscale frames to all 50 seed frames. The frames that were the most similar where said to have the highest probability of being a "good" frame. This required 320,584,700

[S1] `https://www.c-span.org/`

[S2] `https://www.voteview.com/`

[S3] `http://congressionalbills.org/`

Figure S3: Explaining Motion Detection and How the Overhead Shots Were Extracted from C-SPAN Videos

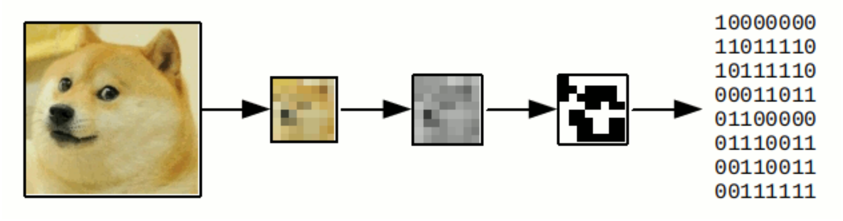


(1) Using ffmpeg extracted frames from 6,526 C-SPAN videos (2009-2014).
(2) Used a perceptual hashing algorithm to create a fingerprint for each frame.
(3) Compared the fingerprints for all frames with fingerprints of 17,170 base frames.
(4) If a frame was sufficiently similar to the base frames, then it was set aside.
(5) Using sequential frames, created 70,710 clips and added meta data to each.
(6) Used frame differencing and the Structural Similarity (SSIM) to calculate the aggregate motion for each clip.
(7) Paired the clips with the corresponding roll call votes to create final dataset.

1	video	frame1	frame2	sim
2	house-vote-105-1-318	house-vote-105-1-318_f158.jpg	house-vote-105-1-318_f159.jpg	0.96414914
3	house-vote-105-1-318	house-vote-105-1-318_f162.jpg	house-vote-105-1-318_f163.jpg	0.96137325
4	house-vote-105-1-318	house-vote-105-1-318_f170.jpg	house-vote-105-1-318_f171.jpg	0.95554648
5	house-vote-105-1-318	house-vote-105-1-318_f171.jpg	house-vote-105-1-318	14
6	house-vote-105-1-318	house-vote-105-1-318_f172.jpg	house-vote-105-1-318	94
7	house-vote-105-1-318	house-vote-105-1-318_f173.jpg	house-vote-105-1-318	29
8	house-vote-105-1-318	house-vote-105-1-318_f174.jpg	house-vote-105-1-318	93
9	house-vote-105-1-318	house-vote-105-1-318_f175.jpg	house-vote-105-1-318	29
10	house-vote-105-1-318	house-vote-105-1-318_f176.jpg	house-vote-105-1-318_f177.jpg	0.96497345

Note: Please see pages S6 – S9 for more details about how the overhead shots were extracted and video motion was detected.

Figure S4: Using a Perceptual Hash to Summarize the Information Found in an Image



Note: This perceptual hash example was borrowed from Essaya (2016). Moving from left to right, the example starts with the full image. In the second step, the image is compressed to a thumbnail which is then converted to grayscale and standardized in the third step. The DCT is used to separate the spectral bands in the fourth step. Once this done, the thumbnail can be expressed as a matrix of 1s and 0s (see fifth step), ultimately representing the most identifiable portions of the image.

pairwise comparisons. To speed up the calculation, I created a fingerprint for each frame using a hashing algorithm. Although there are a number of algorithms one can use, I used one developed by Zauner (2010) commonly known as the "perceptual hash."

Unlike the average hash, a perceptual hash will give similar (or near similar) results for images that have been slightly distorted. This is important for the present application since the videos vary in quality. As shown in Figure S4, the perceptual hash involves the following steps:

1. Convert the image to a thumbnail of 32 by 32 pixels
2. Convert the thumbnail to grayscale
3. Standardize the pixel values
4. Use a two-dimensional Discrete Cosine Transform (DC) to separate the image into spectral sub-bands
5. Take the top left of the resulting DCT to identify the low frequency components that are the most significant
6. If a cell in the resulting 8 × 8 matrix is positive, then record the cell as 1. If the cell is negative, record the cell as 0
7. Flatten the binarized matrix into a vector of 1s and 0s and convert the vector into a 16 bit hexadecimal string

Unlike other image similarity measures, comparing perceptual hashes is less computationally intensive. Even though all 6,411,694 grayscale frames have to be loaded into memory to create the initial hashes, this does not have to be done repeatedly which ultimately makes comparing a very large number of images possible.

Step 4: Finding "good" frames. With these fingerprints in hand, I calculated the degree to which each frame was similar to all 50 seed frames by calculating the hash similarity. For example, let's assume the perceptual hashes for Images A (a10d8ef1c8c87d7a) and B (a10d8ef1dd4623b7) are 50 percent the same. Conversely, the perceptual hashes for Images A (a10d8ef1c8c87d7a) and C (a10d8ef1c8c87d7b) are 93.75 percent the same. This would suggest Image A is more similar to Image C as compared to B. I used essentially the same calculation to score each frame using all 50 seed frames.

Frames which had hash similarity scores in the 95th percentile were then downloaded and used in an unsupervised clustering algorithm to create 4,000 separate groups. A research assistant then reviewed the first five frames within each group to determine whether the group had frames similar to the "ant farm" shot. If it did, then the group was said to be composed of "good" frames. After this process was repeated for all 4,000 groups, I had compiled a list of 17,700 "good" frames.

I used these frames to then score all 6,411,694 frames by comparing each frame to all 17,770 "good" frames using the hash similarity. Frames with higher hash similarity scores were said to have a greater likelihood of being the "ant farm" shot. Ultimately, I used a high performance computing cluster to calculate the 113,486,983,800 hash similarity scores. From start to finish, the calculation took around a day.

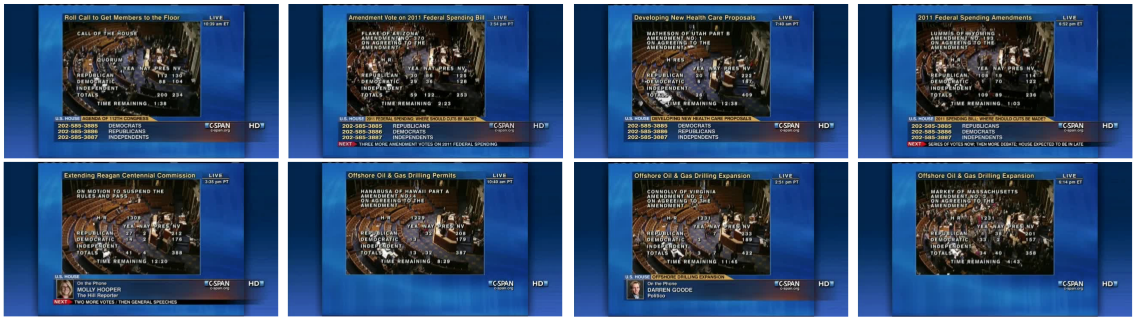
Frames were said to be similar enough to one of 17,700 "good" frames when they shared 10 of 16 hexadecimal characters. If a frame was sufficiently similar to at least one of the "good" frames, then it was included in a dataset referred to as "PHash = 0." The "PHash = 50" and "PHash = 100" datasets include frames that were sufficiently similar to at least 50 and 100 "good frames," respectively. I report the results from the "PHash = 0" dataset in the main text, but results from the other two datasets are included in Section S3.2 of the SI as robustness checks.

Figure S5 demonstrates the methodology introduced above yields reasonable results. In Panel A, I provide the four frames that are *most* similar to every other frame from videos with a scaled SSIM score between 0 and half a standard deviation above the mean. In Panel B, I do the same for videos with a scaled SSIM score between half a standard deviation below the mean and 0. Both panels are remarkably similar to the "ant farm" shot found in Figure S1 which implies hash similarity can be used to effectively differentiate between "good" and "bad" frames.

S1.4 Measuring Aggregate Motion

Originally developed by Wang et al. (2004), the Structural Similarity Index (SSIM) has a number of desirable properties, most notably its ability to determine whether images are perceptually similarly. Please refer to Hu et al. (2004) and Yilmaz, Javed and Shah (2006) for information on frame differencing and video analysis more broadly. For an example of

Figure S5: Visualizing the Data Using the Most Similar Frames



(a) $0.50 >$ Structural Similarity > 0

(b) $0 >$ Structural Similarity > -0.50

Roll Call to Get Members to the Floor
Amendment Vote on 2011 Federal Spending Bill
Developing New Health Care Proposals
2011 Federal Spending Amendments
Extending Reagan Centennial Commission
Offshore Oil & Gas Drilling Permits
Offshore Oil & Gas Drilling Expansion
Offshore Oil & Gas Drilling Expansion

Note: In Panel A, I provide the four frames that are *most* similar to every other frame from videos with a scaled SSIM score between 0 and half a standard deviation above the mean. In Panel B, I do the same for videos with a scaled SSIM score between half a standard deviation below the mean and 0.

SSIM being used for motion detection please see Seshadrinathan and Bovik (2007).

Figure S6 shows why SSIM is preferable to using the Mean Squared Error (MSE) for frame differencing. In Panel A, a baseline image of Albert Einstein is provided. The Mean Squared Error (MSE) is the same in Panels B-D even though the baseline image has been noticeably altered. Wang and Bovik (2009) argues this is because the MSE assumes:

1. Image similarity is independent of temporal or spatial relationships, meaning randomly re-ordered images can yield the same MSE.

2. Image similarity is independent of any relationship between the original signal and the error signal. For a given error signal, the MSE remains unchanged regardless of the base image.

3. Image similarity is independent of the sign of within sample error (e.g., error is generally positive or negative).

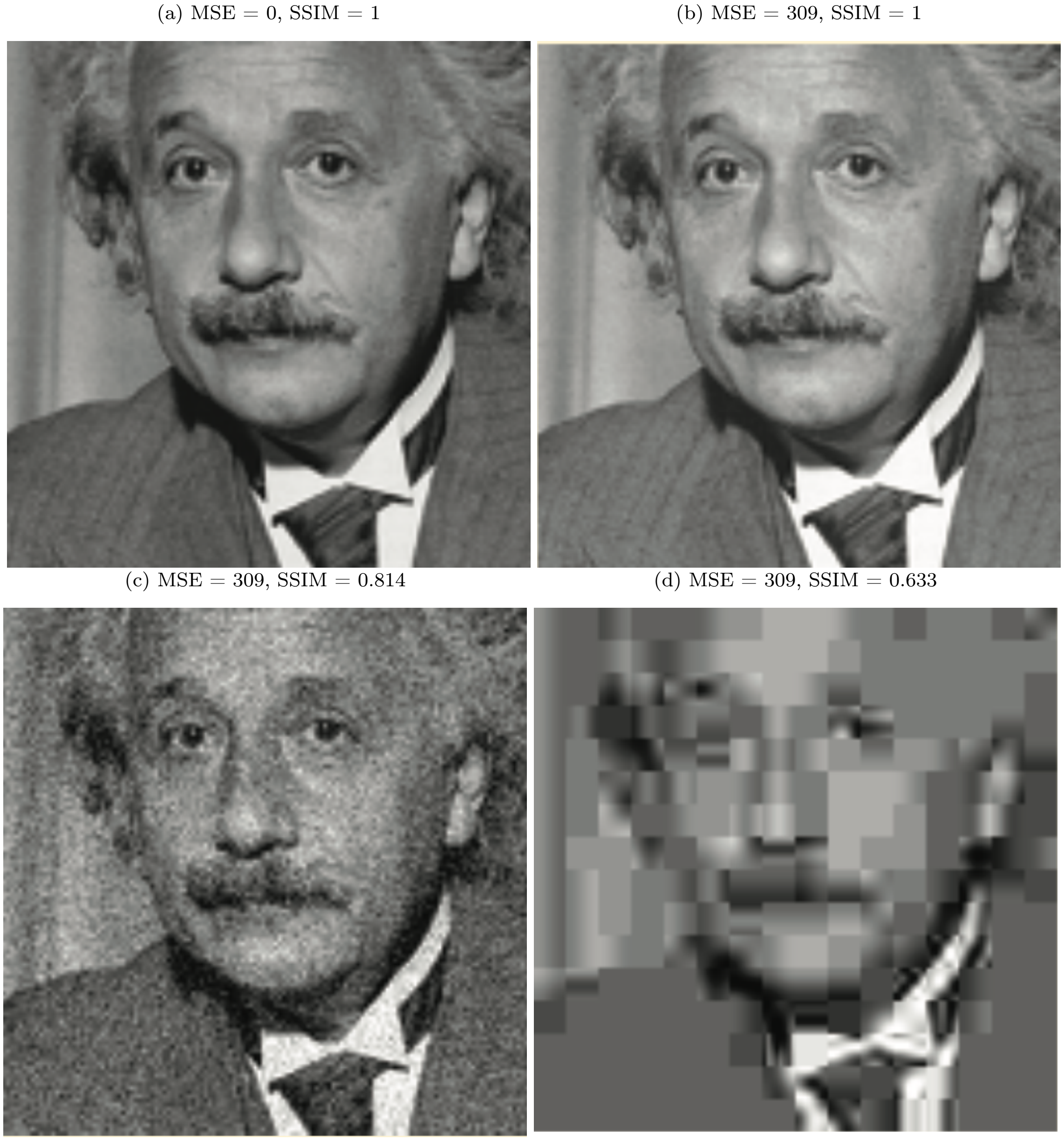
4. Image similarity is independent of the types of signals sampled. Regardless of which portion of the image is selected, the MSE should be the same.

Undoubtedly, these are very strong assumptions which often produce disparate results similar to those shown in Figure S6. The SSIM is better able to capture the subtle changes in the Albert Einstein images because it takes into consideration the luminance, contrast, and structural similarity of each image.

If $\boldsymbol{x}$ and $\boldsymbol{y}$ are the pixel matrices associated with two images, then the SSIM can be defined as:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (1)$$

Figure S6: Using the Structural Similarity Index (SSIM) to Quantify Image Differences



(a) MSE = 0, SSIM = 1

(b) MSE = 309, SSIM = 1

(c) MSE = 309, SSIM = 0.814

(d) MSE = 309, SSIM = 0.633

Note: This example was borrowed from Wang and Bovik (2009). The original image is provided in Panel A. A luminance shift is applied in Panel B. Gaussian noise is added in Panel C and the image is compressed in Panel D. I include the MSE and SSIM when each image is compared to Panel A. In Panels B-D, the MSE remains the same despite the images being noticeably different.

,where μ_x and μ_y are the means of x and y, respectively. Similarly, σ_x and σ_y are the standard deviations of x and y, leaving σ_{xy} as the cross correlation of x and y. C_1 and C_2 are "small positive constants that stabilize each term, so that near-zero sample means, variances, or correlations do not lead to numerical instability" (Wang and Bovik, 2009, 106). In `Python` these are defined as follows:

$$\begin{aligned} C_1 &= (0.01 \times (max_{\boldsymbol{x},\boldsymbol{y}} - min_{\boldsymbol{x},\boldsymbol{y}}))^2 \\ C_2 &= (0.03 \times (max_{\boldsymbol{x},\boldsymbol{y}} - min_{\boldsymbol{x},\boldsymbol{y}}))^2 \end{aligned} \tag{2}$$

where $max_{\boldsymbol{x},\boldsymbol{y}}$ and $min_{\boldsymbol{x},\boldsymbol{y}}$ are the maximum and minimum pixel values across both matrices. If either matrix has any white, then $max_{\boldsymbol{x},\boldsymbol{y}}$ will equal 1. Similarly, any black will set $min_{\boldsymbol{x},\boldsymbol{y}}$ to zero, making C_1 and C_2 0.0001 and 0.0003, respectively.[S4]

Using the "ant farm" frames recovered using the process described above, I used the average SSIM to estimate the degree to which objects were moving in each video. This required comparing sequences of frames which I call "clips." On average, each video contained 17 clips, meaning within each video there were approximately 17 uninterrupted sequence of "ant farm" frames. The aggregate motion of each video was estimated using the pairwise structural similarity between each frame within each clip.

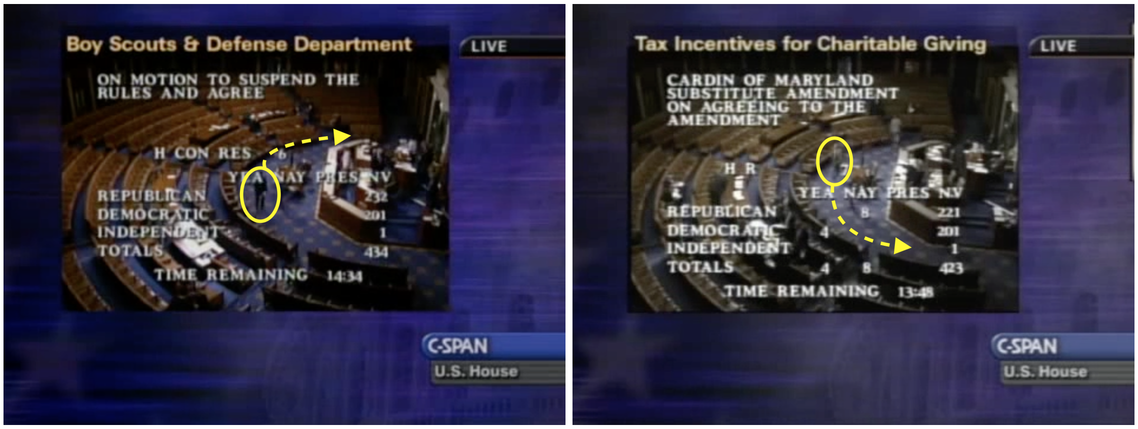
For example, let's assume a video had two clips: a and b. If the first clip (a) had three frames (a_1, a_2, and a_3) and the second clip (b) had four frames (b_1, b_2, b_3, b_4), then the average SSIM for this video would be .80 assuming the following:

$SSIM(a_1, a_2) = .65$
$SSIM(a_2, a_3) = .90$
$SSIM(b_1, b_2) = .70$
$SSIM(b_2, b_3) = .80$
$SSIM(b_3, b_4) = .95$

Videos with frames that are similar tend to produce less motion. For example, the first clip (a) has an average SSIM of 0.78, whereas the second clip (b) has an average SSIM of 0.82, suggesting the frames in the second clip were more similar to one another than the frames in the first. Ultimately, this implies the first clip (a) has slightly more motion since the frames tend to change more from one frame to the next. To help interpret the measure, I scaled the average SSIM to standard deviations above and below the mean – with positive values implying *less* motion.

[S4] It is also important to note the SSIM works quite well even when $C_1 = C_2 = 0$ (see Wang and Bovik, 2009, 106)

Figure S7: Examples of MCs Walking in the Well of the House



(a) Walking South to North

(b) Walking North to South

Note: In Panel A, I show an MC walking across the well of the House from South to North. In Panel B, a different MC is walking North to South.

S2 Validation Exercises

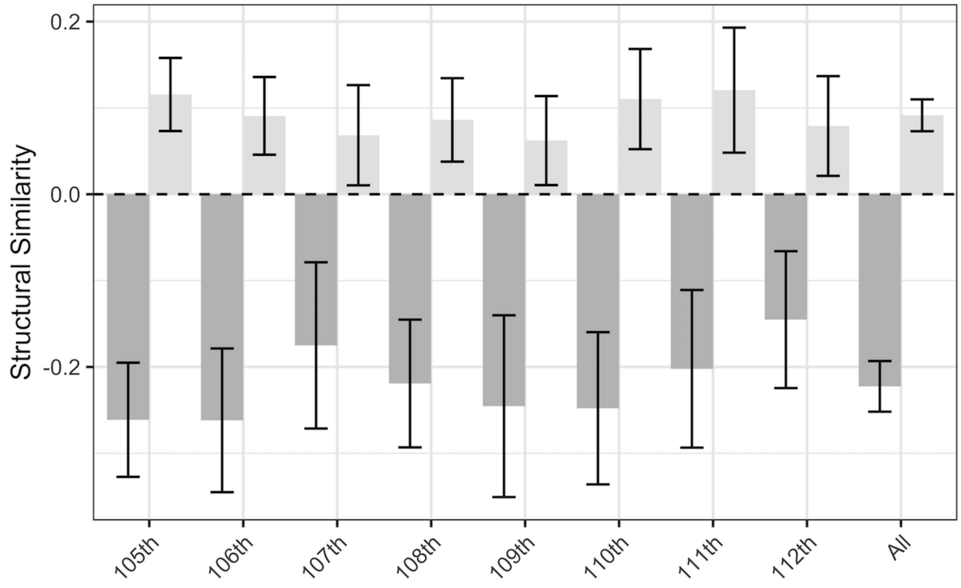
S2.1 Validation #1: Walking in the Well of the House

In my first validation exercise, I determined whether the average structural similarity is correlated with MCs walking in the House. To do so, I randomly selected 500 clips from the 70,717 clips I compiled for this study. The sample was then restricted to only videos that included at least one MC walking in the well. Once these 405 videos were identified, I then manually coded each frame for whether an MC was walking in the well of the House. The result was a binary vector equal to the 16,628 frames in my reduced sample.

Figure S7 gives an example of two frames in which an MC was walking in the well of the House. In Panel A, an MC is walking from South to North, whereas in Panel B an MC is walking in the opposite direction. In each instance, the frame would be coded as 1 since an MC is walking in the well of the House. The MC cannot be identified in either frame which is why I used several validation exercises to determine whether bipartisan interactions are more likely in videos with less motion.

Figure S8 reports the results from this validation exercise. Here, I standardized the structural similarity for each frame using the mean and standard deviation for each video clip. This resulted in a measure that was positive when the frame had less motion. This standardization approach was used to account for some slight variations in the "ant farm" shot from one video to the next. Once this standardized variable was created, I then compared the average of frames in which MCs were walking to the well (1) versus those in which the MCs were not walking in the well of the House (0).

Figure S8: Structural Similarity is *Lower* (Implying *More* Motion) When MCs are Walking in the Well of the House



Note: The dark gray box (■) indicates at least one MC was walking in the well of the House, whereas the light gray box (□) indicates no MCs were walking in the well. Average structural similarity is shown in the y-axis. Positive values imply the frames were more similar, implying *less* motion. Vertical lines represent 95 percent confidence intervals.

The y-axis shows the average structural similarity across all sampled clips in which MCs were (see light grey bars) or were not (see dark grey bars) walking in the well of the House. Regardless of the Congress, Figure S8 shows frames with MCs walking in the well of the House have significantly (p $<$ 0.001) *less* structural similarity implying there is significantly *more* motion. This suggests videos with more motion are more likely to have Democrats and Republicans walking in the well of the House.

In instances where MCs were moving in other sections of the House, the variable was still recorded as zero which suggests I am reporting a conservative estimate. If I were to restrict the analysis to only frames in which MCs were either standing or walking in the well of the House, I imagine the difference in motion would be even starker. With that said, the difference between the motion in frames where MCs were walking in the well versus those where they were not was highly significant ($p < 0.001$) which suggests video motion can be used to reasonably capture MCs walking in the well of the U. S. House of Representatives.

S2.2 Validation #2: Crossing the Aisle

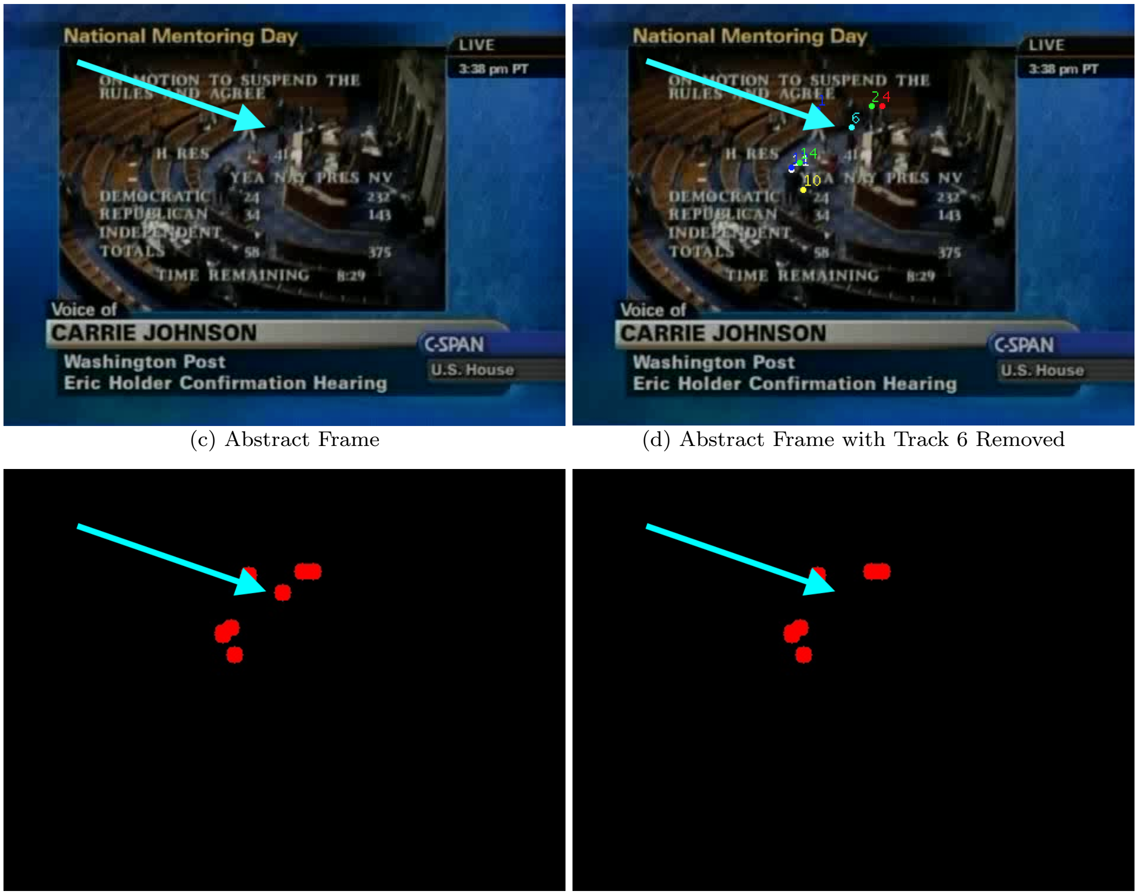
In the second validation exercise, I determine whether video motion increases when MCs literally cross the aisle. To do so, 100 clips were randomly sampled from the 405 clips outlined above. An undergrad research assistant was then asked to manually track every MC who entered the well of the House using the *Fiji* distribution of *ImageJ*.[S5] If an MC was already in the well of the House when the clip began, that MC was also tracked.

After the manually tracking was complete, abstract representations of each video were then create. Here, I replaced all MCs with red dots using the tracking data produced by my undergraduate research assistant. Each MC was then sequentially extracted and new clips were created. Ultimately, this made it possible to determine the contribution of each MC's movement to the aggregate motion of the clip. If crossing the aisle produces more aggregate motion, then MCs who literally walk from one side of the room to the other should significantly decrease the overall video motion when they are removed from the video since there movement generally produced more motion.

Figure S9 provides an example of how this measure is calculated. In Panel A, I show a frame from one of the randomly selected video. The light blue arrow is pointing to the MC that will eventually be extracted from the video. Panel B shows the manually tracked MCs with an arrow pointing to the MC labeled six. The tracking information is then used to create an abstract representation of the frame which is shown in Panel C. In this version, the rest of the video information and the labels are removed. What is left is a red dot for each tracked MC. In Panel D, I create the same abstract frame, but I do not include the MC labeled six.

[S5] `https://fiji.sc/`

Figure S9: An Example of the Second Validation Exercise



(a) Original Frame

(b) Tracked Frame

(c) Abstract Frame

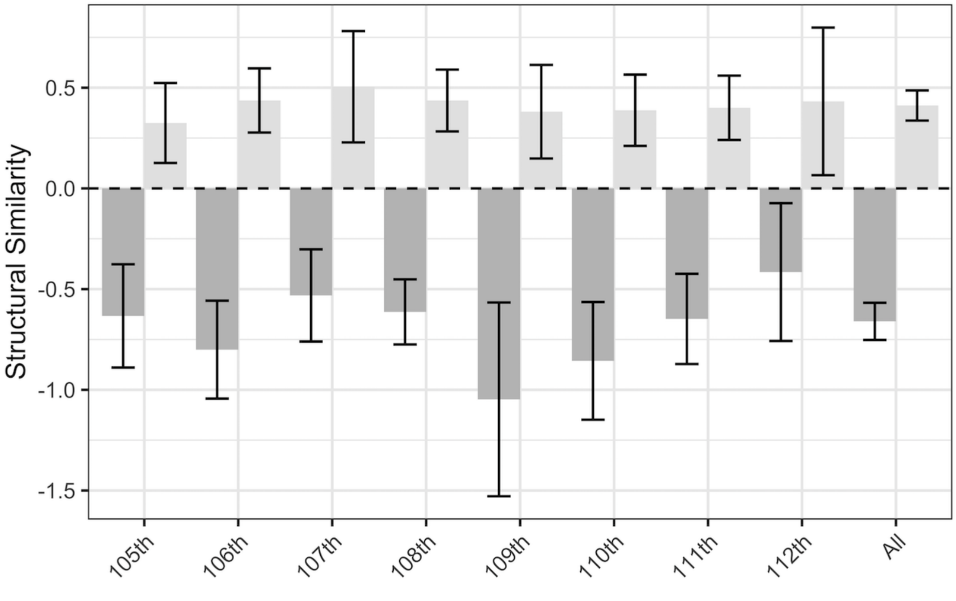
(d) Abstract Frame with Track 6 Removed

Note: In Panel A, I show a frame from one of the 100 randomly selected clips. In Panel B, I show the manually tracked MCs. In Panel C, I show the abstract representation of each frame. Finally, Panel D shows the same abstract representation without the MC labelled six. The light blue arrows are meant to help the reader to interpret the figure. They were not included in the actual video analysis.

From this point, the analysis is identical to what was presented in the main text. After the MC labeled six is removed from the video, the aggregate motion of the video without that MC is calculated. The result is a measure of that MC's contribution to the overall motion in the video. If the MC labeled six moves a lot, then the overall motion of the video will decline when that MC is removed. If the same MC never moves, then the overall motion of the video will remain the same. Ultimately, this provides a useful measure of the influence of each MC on the aggregate measure I used in the main text.

The final step in this validation exercise is to determine whether the MC crossed the aisle. Here, I looked for MCs following the patterns outlined in Figure S7. If the MC walked across the aisle from either South to North or North to South, then the MC was said to "cross the aisle." All other MCs are recorded as zeros resulting in a binary vector equal to the number of MCs who entered the well of the House in the 100 randomly sampled videos outlined above.

Figure S10: Structural Similarity is *Lower* (Implying *More* Motion) When MCs are Crossing the Aisle



Note: The dark gray box (■) indicates the MC crossed the aisle, whereas the light gray box (□) indicates the MC did not cross the aisle. Average structural similarity is shown in the y-axis. Positive values imply the frames were more similar, implying *less* motion. Vertical lines represent 95 percent confidence intervals.

Figure S10 reports the results from this validation exercise. However, to make these results comparable to Figure S8, I inverted the scale. If an MC's walking pattern contributes considerably to the overall motion of the video, when he/she is removed the frames will become *more* similar to one another. This is because the MC's walking pattern is the main reason why the frames in the original video were dissimilar since he/she is moving considerably causing the frames to be less similar. Given that, MCs who crossed the aisle should *increase* the structural similarity of the videos when they are removed since the videos now have *less* overall motion.

Once the scale was inverted, the results in Figure S10 are remarkably similar to those reported in the previous subsection. Regardless of the Congress, when MCs cross the aisle they produce significantly ($p < 0.001$) *more* motion which suggests videos with more motion are also more likely to have Democrats and Republicans walking across the aisle.[S6] Since I am unable to manually track and code every MC in the 70,717 C-SPAN clips, it difficult to say how many Democrats and Republicans talk to one another after each floor vote. However, these two validation exercises show bipartisan interactions are more likely to be found in videos with more motion.

S2.3 Validation #3: Agent-Based Model

To gain more traction on the causal question, I created an agent-based model which simulates the social interactions immediately after floor votes (for review see Zhou et al., 2010). Here, the environment is a simple 250 by 250 matrix. Agents (112 "Republicans" and 96 "Democrats") are then randomly assigned two "vision" parameters, one of which allowed the agent to look north and south while the other allowed the agent to look east and west. These "vision" parameters were randomly drawn from a uniform distribution that ranged from 1 to 200, meaning at one extreme agents could only look one space around them while on the other extreme they could look almost across the room. Even though this parameter can be interpreted as some MCs being able to see further than others, a more reasonable interpretation is the willingness to seek out a discussion partner. Agents with greater "vision" are willing to expend more effort to find others to talk to, whereas the inverse is true for agents with less "vision."

After this, each agent was assigned a "movement" parameter, which was randomly drawn from a uniform distribution which ranging from .50 to 1. This variable captures the degree to which an agent is likely to move. Although I wanted to make it more likely than not that an agent moved, I also wanted to allow some agents to be less willing to budge as compared to others. Similarly, I made some agents "faster" than others, meaning at any given time step some could move more spaces than others. This parameter was also set using a random uniform distribution (min = 1, max = 10). Finally, although the goal of this simulation is to mimic social interactions, some agents may be social butterflies, meaning instead of just talking to one person they want to mix and mingle. This was captured using a variable randomly drawn from a uniform distribution which ranged from 0 to .25, meaning that on average agents are not going to jump from one agent to another, but some may.
In the initial time step of the polarization simulation,[S7] each agent first decides whether to

[S6] Unfortunately, the random sample only included a single video from the 111^{th} Congress. Given that, I imputed these results by averaging across the 110^{th} and 112^{th} Congresses.

[S7] The polarization simulation can be found here: `https://youtu.be/kNBmFpqdBkw` move. If the agent chooses to move, the agent then looks north, south, east, and west for agents around them. The degree to which they can see other agents is constricted by their "vision." Once they obtain a list of potential targets, they only select targets that are from their own party, meaning Democratic agents only select Democrats and Republican agents only select Republicans. The bipartisan simulation is essentially identical except instead of trying to find targets from their own party each agent is trying to find targets from the opposition,[S8] meaning Democratic agents seek out Republicans and Republican agents seek out Democrats.

Regardless of the simulation, once potential targets are selected the agent determines which is closer, then moves towards that target. This motion is first determined by taking a number of steps north, south, east, and west equal to each agent's "speed." Once these potential moves are calculated, the agent selects the move that minimizes the difference between it and the partisan target. After this move is made, the agent then records its position and the id of the target. In some instances, the agent will be unable to find a target. When this happens the agent randomly moves (equal to the agent's "speed) either north, south, east, or west.

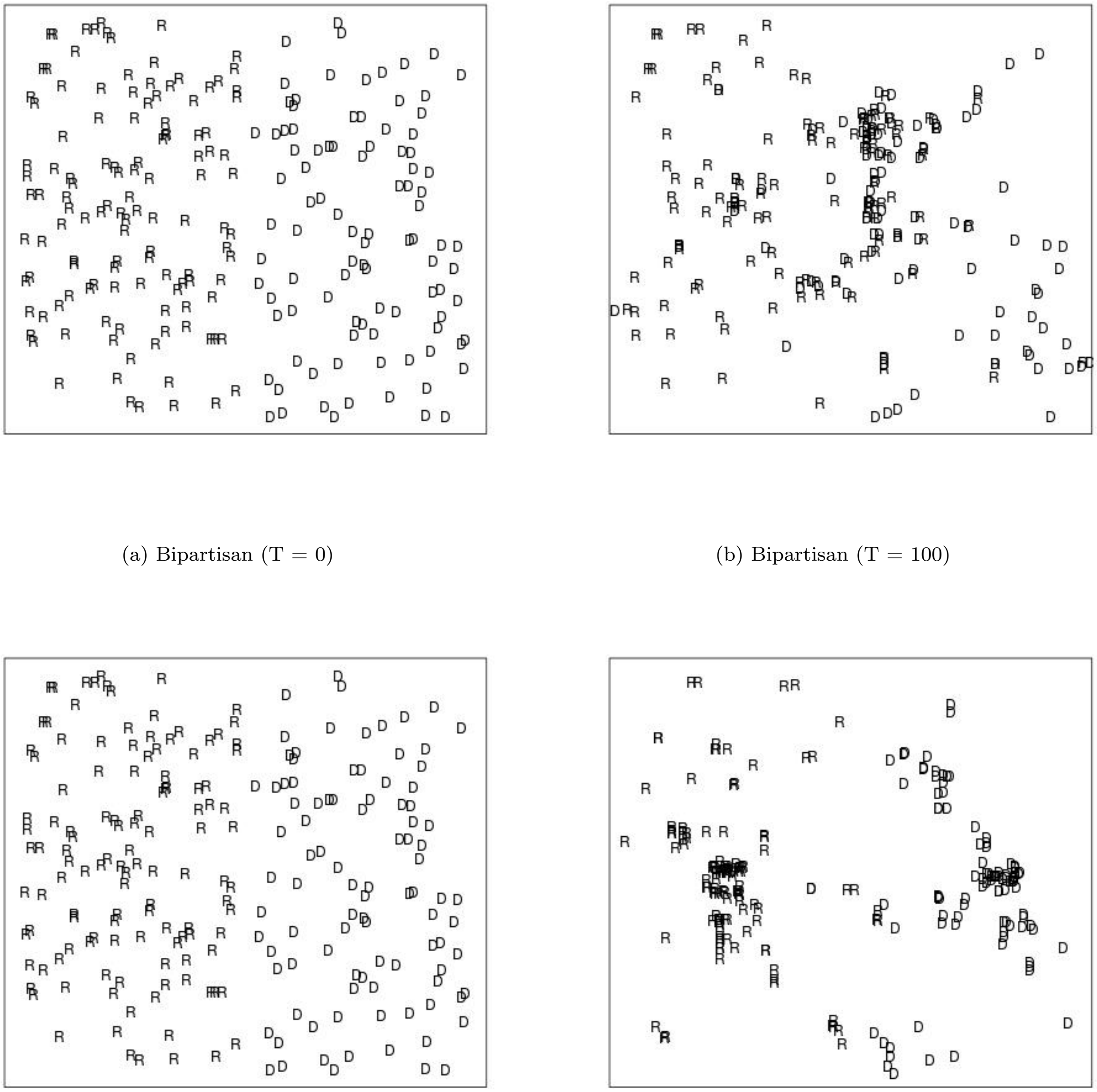
From this point, the simulation continues in a similar fashion in subsequent time steps, with two caveats. First, if an agent has already found a target, then the agent proceeds to move towards that target. This was done because I assume that agents are seeking out their friends in the legislature. Generally, these "friendships" are stable, meaning they tend to select one friend and stick with that selection. Thus, if the agent does not have a partisan target, then the agent follows the process outlined above in order to find one. Second, in subsequent time steps the agent can decide if they want to find a new target. If they do, then their current partisan target is removed from their memory and they find a new one using the process outlined in the previous paragraph. Of course all of this assumes that the agent has selected to move in the given time step. If the agent has not selected to move, then the agent stays put.

Figure S11 shows the initial and final positions of the agents after 100 time steps. The first thing to note is how the Democratic (represented by a "D") and Republican (represented by a "R") agents are positioned on either side of the board. In the House there are no assigned seats, but are, by tradition, divided by party, with Democrats sitting to the Speaker's right and the Republicans sitting to the Speaker's left. As you can see, the same distribution of agents appear for both the bipartisan and polarization simulations. This is because each simulation uses the same initial conditions.

Figure S12 compares the average change in pixel intensity using each frame from the bipartisan and polarization simulations directly. Positive values imply more change exists in the former as compared to the latter. For example, at the 20th time step the average change

[S8] The bipartisanship simulation can be found here: `https://youtu.be/Ot1xerXV9qw`

Figure S11: Images of Initial and Final States of Each Agent-Based Simulation



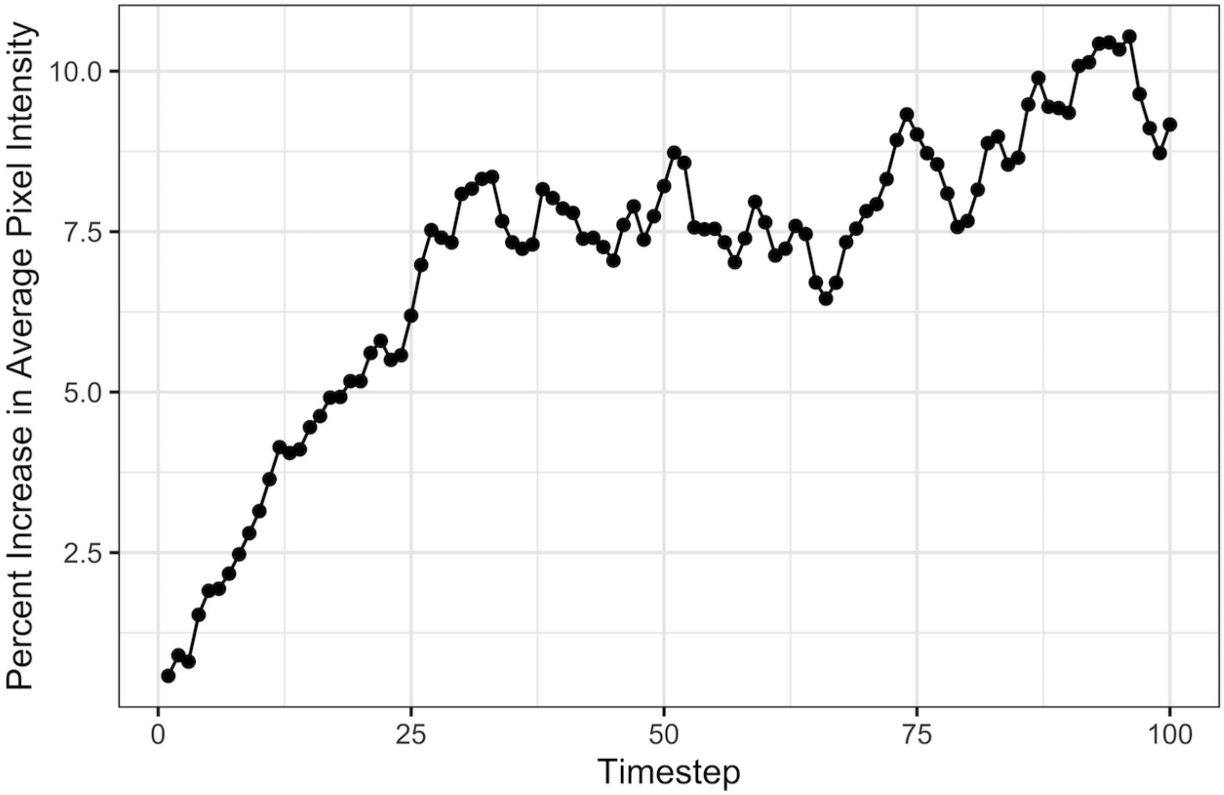
(a) Bipartisan (T = 0)

(b) Bipartisan (T = 100)

(c) Polarization (T = 0)

(d) Polarization (T = 100)

Figure S12: Results from an Agent-Based Model of Bipartisan versus Polarized Social Interactions on the House Floor



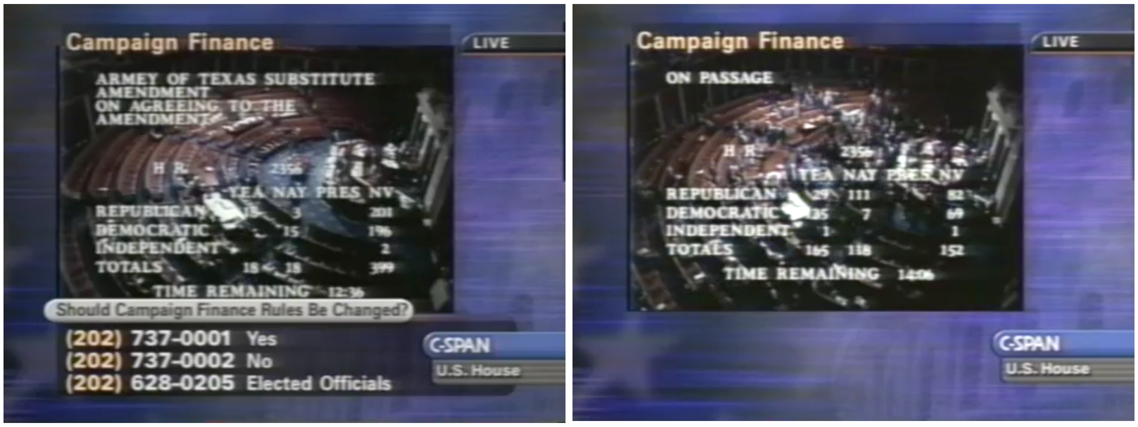
in pixel intensity was about four percent higher in the bipartisan simulation than the average change in pixel intensity for the polarization simulation. With this in mind, when comparing the bipartisan to polarization simulations, the former, on average, has more changes in pixel intensity, suggesting in the bipartisan simulation more movement is present.

In Sections S2.1 and S2.2, I show aggregate video motion is associated with MCs walking in the well of the House and literally crossing the aisle. In the above agent-based model, I simulate the types of social interactions that occur after floor votes and find strong evidence that video motion is highly correlated with bipartisan social interactions. Collectively, these results demonstrate that the measure I introduce in this study is reasonably correlated with Democrats and Republicans speaking (or not speaking) to one another on the floor of the U.S. House of Representatives.

S2.4 Validation #4: Bipartisan Campaign Reform Act (BCRA)

Finally, videos associated with the Bipartisan Campaign Reform Act (BCRA) are used to not only provide some additional validation, but also to give readers a sense of what video motion looks like. When this bill was debate in 2002, several amendments were offered to derail the bipartisan effort (see Figure S13, Panel A). In each instance, not only did the majority of Democrats vote against the majority of Republicans, but they also turned to the House floor to demonstrate their commitment to the bipartisan effort. This culminated with nearly every representative pouring onto the House floor to congratulate one another after the legislation was passed (see Figure S13, Panel B).

Figure S13: Overhead Shot of Members of Congress Mingling after the First and Last Roll-Call Vote on the Bipartisan Campaign Reform Act of 2002 (BCRA)

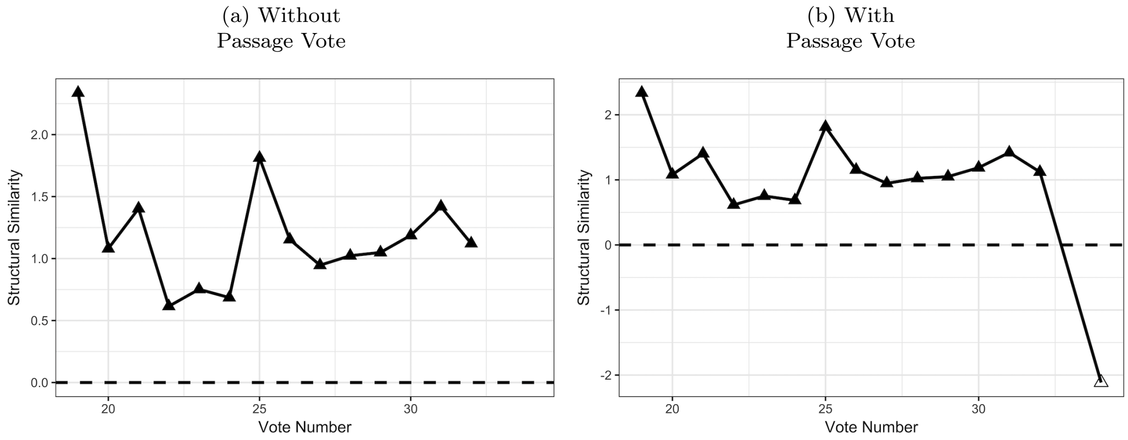

(a) First Vote

(b) Passage Vote

Note: In Panel A, I show a frame of Democrats and Republicans talking after Dick Armey's (R-TX) substitute amendment which attempted to derail BCRA. In Panel B, I show a frame of Democrats and Republicans pouring onto the House floor to celebrate the passage of BCRA.

Figure S14 shows this ebb and flow can be captured using the measure introduced in this study. Beginning with Panel A, we can see videos associated with the first vote did not contain a lot of motion. From there, the second and third videos contained more and more motion, implying an increasing number of bipartisan interactions between votes. More specifically, the first and third votes were on substitute amendments, but the first amendment was proposed by Dick Armey (R-TX) whereas the third was proposed by Chis Shay (R-CT) who sponsored BCRA in the House. Representative Armey's amendment attempted to derail BCRA by immediately banning the use of soft money, whereas Representative Shay's amendment "was nearly identical to the bill" and represented a compromise reached by Democrats and Republicans where the soft money provisions would take effect after the upcoming congressional election.

Figure S14: Assessing the Number of Bipartisan Social Interactions During the BCRA Debate


(a) Without Passage Vote

(b) With Passage Vote

Note: Panel A plots the scaled SSIM for each BCRA vote without including the passage vote. Panel B plots the same time series with the passage vote. An image of the first vote and the passage vote can be found in Figure S13. Positive values imply the frames are more similar to one another implying there is *less* motion.

After Shay's amendment, the video motion remains stable until it decreases dramatically after the sixth roll call vote. This vote was on an amendment proposed by J.C. Watts (R-OK) that exempt "communication pertaining to civil rights and issues affecting minorities" from BCRA. This amendment was particularly divisive because it was proposed in response to an amendment offered by Chip Pickering (R-MS) which attempted to create an "exemption for communications pertaining to the Second Amendment of the Constitution." After these votes, the change in SSIM suggests there were likely fewer bipartisan social interactions, especially after the Watts amendment.

BCRA is often offered as one of the few examples of successful bipartisan legislation. Panel B shows Democrats and Republicans may have recognized the significance of their achievement. Not only does the video motion increase after BCRA passed the House (as indicated by the *decrease* in SSIM), but it changes the scale of the graph so much that it makes the votes on the Armey, Shay, Watts, and Pickering amendments pale in comparison. This is due to a large number of Democrats and Republicans literally crossing the aisle to speak with one another after the passage vote.

Even though it is impossible to say definitively how much motion is produced when Democrats and Republicans cross the aisle, Figure 3 in the main text and Figure S14 show increased video motion is associated with a more bipartisan political environment.

S3 Robustness Checks

S3.1 Potential Outliers

To address concerns that my results are being driven by videos with an extraordinary amount of motion, I re-estimated all models in the main text eliminating potential outliers. To do so, I first use a very restrictive definition of what constitutes an outlier: any video with a structural similarity score more than ±2 standard deviations away from the mean. Ultimately, the results we remain the same when the data is restricted in this way.

More specifically, in Table S2, the models are the same as those reported in Table 1 in the main text, but they have been estimated excluding video outliers. Regardless of the model, `Structural Similarity` is positive and statistically significant implying videos which have *less* motion are more likely to be associated with future party votes. This provides strong evidence that the results reported in the main text cannot be attributed to a handful of extreme videos.

Table S3 reports a similar analysis, but instead of defining an outlier as any video with a structural similarity score more than ±3 standard deviations away from the mean, the definition is expanded to ±2 standard deviations. The results remain the same when the latter definition is used. Indeed, `Structural Similarity` is always positive and statistically significant in Tables S2 and S3, suggesting the results outlined in Table 1 cannot be attributed to extreme cases that deviate from the general trend in the data.

Regardless of the model, excluding extreme values does not affect the substantive results. Indeed, `Structural Similarity` is always statistically significant and in the same direction as the coefficients reported in Table 1. This provides strong evidence that the results reported in the main text cannot be attributed to a handful of influential observations. The robustness of my results is further underlined in the next section. In that section, I re-estimate all the models using different data configurations of the "ant farm" data. Similar to the models reported in this section, the coefficients reported in Table 1 are the same suggesting my main results are robust to a handful of extreme observations.

Table S2: When MC's Cross The Aisle, Future Party Votes Are Less Likely To Occur ("PHash = 0" with No Outliers 3 SD)

	Dependent variable:					
	Future Party Votes					
	(1)			*(2)*		
Variable	$\hat{\beta}$	$\text{SE}_{\hat{\beta}}$	95% CI	$\hat{\beta}$	$\text{SE}_{\hat{\beta}}$	95% CI
Constant	0.198	0.055	$[0.091, 0.306]$	−0.530	0.144	$[-0.812, -0.248]$
Structural Similarity	0.070	0.029	$[0.012, 0.128]$	0.045	0.025	$[-0.005, 0.095]$
Previous Party Votes				0.497	0.093	$[0.314, 0.679]$
Passage Vote				0.012	0.049	$[-0.084, 0.107]$
Amendment Vote				0.486	0.060	$[0.367, 0.604]$
Total Not Voting				−0.020	0.003	$[-0.026, -0.014]$
|Sponsor Ideology|				0.292	0.164	$[-0.030, 0.614]$
Sponsor Seniority				−0.009	0.006	$[-0.021, 0.002]$
Sponsor Party Leader				−0.124	0.163	$[-0.444, 0.195]$
Election Year				1.163	0.098	$[0.971, 1.355]$
N	3,567			3,567		
Log Likelihood	−3532.130			−3155.720		
AIC	7084.260			6347.439		

Note: In these models, any video with `Structural Similarity` more than ±3 standard deviations away from the mean are excluded. `Structural Similarity` is described on page 6 in the main text and pages S9–S12 in the SI. Positive values imply *less* video motion. Unit of analysis is a given floor vote. Since some House bills and resolutions have several votes, standard errors are clustered around each piece of legislation (e.g., `HR` 820). 95% confidence intervals are also reported. All models include Congress fixed-effects and were estimated using the **tobit** function in **Stata** (v15).

Table S3: When MC's Cross The Aisle, Future Party Votes Are Less Likely To Occur ("PHash = 0" with No Outliers 2 SD)

			Dependent variable:			
			Future Party Votes			
		(1)			*(2)*	
Variable	$\hat{\beta}$	$\text{SE}_{\hat{\beta}}$	95% CI	$\hat{\beta}$	$\text{SE}_{\hat{\beta}}$	95% CI
Constant	0.196	0.055	$[0.088, 0.303]$	-0.549	0.144	$[-0.830, -0.267]$
Structural Similarity	0.086	0.031	$[0.025, 0.146]$	0.065	0.027	$[0.012, 0.117]$
Previous Party Votes				0.493	0.093	$[0.312, 0.675]$
Passage Vote				0.008	0.049	$[-0.088, 0.104]$
Amendment Vote				0.486	0.060	$[0.368, 0.603]$
Total Not Voting				-0.020	0.003	$[-0.026, -0.014]$
|Sponsor Ideology|				0.317	0.161	$[0.002, 0.633]$
Sponsor Seniority				-0.009	0.006	$[-0.020, 0.002]$
Sponsor Party Leader				-0.111	0.163	$[-0.430, 0.208]$
Election Year				1.159	0.097	$[0.968, 1.349]$
N	3,502			3,502		
Log Likelihood	-3466.882			-3097.322		
AIC	6953.764			6230.644		

Note: In these models, any video with `Structural Similarity` more than ±2 standard deviations away from the mean are excluded. `Structural Similarity` is described on page 6 in the main text and pages S9–S12 in the SI. Positive values imply *less* video motion. Unit of analysis is a given floor vote. Since some House bills and resolutions have several votes, standard errors are clustered around each piece of legislation (e.g., `HR` 820). 95% confidence intervals are also reported. All models include Congress fixed-effects and were estimated using the **tobit** function in **Stata** (v15).

Table S4: Are Democrats and Republicans Less Willing to Cross the Aisle After Party Votes? ("PHash = 50")

	Dependent variable:			
	Party Vote		Future Party Votes	
	Logistic Regression		*Tobit Regression*	
	(1)	(2)	(3)	(4)
Constant	0.372	0.852	0.054	0.017
	(0.051)	(0.172)	(0.022)	(0.051)
Structural Similarity	0.180	0.210	0.041	0.044
	(0.048)	(0.050)	(0.015)	(0.014)
Passage Vote		0.372		0.224
		(0.129)		(0.037)
Amendment Vote		0.615		0.251
		(0.112)		(0.036)
Total Not Voting		−0.036		−0.004
		(0.006)		(0.002)
|Sponsor Ideology|		−1.324		−0.348
		(0.546)		(0.171)
Sponsor Seniority		−0.016		0.001
		(0.012)		(0.004)
Sponsor Party Leader		1.059		0.081
		(0.382)		(0.093)
Election Year		0.041		−0.060
		(0.114)		(0.046)
N	2,213	2,195	2,213	2,195
Log Lik	−1,487.078	−1,402.598	−1,734.023	−1,650.249
AIC	2,978.155	2,823.196	3,474.047	3,320.498

Note: These models are identical to the models reported in Table 1 except frames had to be sufficiently similar to at least 50 of the "good" frames to be used in the analysis. Dependent variable and the model used are reported above each column. Video motion is captured in the variable labeled "Structural Similarity." Positive values imply the frames within the video are more similar to one another, implying there is *less* motion. The unit of analysis is a given floor vote. The data has also been restricted to votes on House bills and resolutions. Standard errors clustered around the issue and day are reported in parentheses.

S3.2 Varying Number of "Good" Frames

In order to identify the "ant farm" shot, I identify 17,700 "good" frames. I then scored all 6,411,694 frames from the larger video corpus using the hash similarity. Frames with higher hash similarity scores were said to have a greater likelihood of being the "ant farm" shot. Frames were said to be similar enough to one of 17,700 "good" frames when they shared 10 of 16 hexadecimal characters. If a frame was sufficiently similar to at least one of the "good" frames, then it was included in a dataset referred to as "PHash = 0." The results reported in Table 1 use this dataset.

In the tables below, I re-estimate all of the results using two datasets. The "PHash = 50" dataset includes frames that were sufficiently similar to at least 50 of the "good" frames. The "PHash = 100" data set is identical, but frames had to be sufficiently similar to at least 100 of the "good frames. I include these results to demonstrate the results in the main text cannot be attributed to the criteria I used to subset the data. Indeed, the results are identical regardless of whether "PHash = 0," "PHash = 50," or "PHash = 100" are used. This suggests the results presented in the main text are extremely robust.

More specifically, in Table S4 I report the results only including frames which were sufficiently similar to at least 50 of the "good frames." In the first two columns, I report the first two models from Table 1 in the main text using the modified data. In the last two columns, I do the same for Models 3 and 4. Regardless of the model, `Structural Similarity` is positive and statistically significant implying videos which have *less* motion are more likely to be associated with either current (see Models 1 and 2) or future (see Models 3 and 4) party votes. This provides strong evidence that the results reported in the main text cannot be easily attributed to the way the base data was extracted from the larger C-SPAN video corpus.

I find the same results in Table S5. Here, I only include frames which were sufficiently similar to at least 100 of the "good frames." Similar to the previous table, the first two columns include Models 1 and 2 from Table 1 in the main text, whereas Models 3 and 3 are reported in the last two columns. In all four models, I used the more restricted data. Regardless of the model, `Structural Similarity` is positive and statistically significant suggesting my main results cannot be easily attributed to the way I constructed the base data. Indeed, the results seem to also be robust to various data configurations, such as the exclusion of potential outliers. This again underlines the robustness of the results reported in the main text.

S3.3 Controlling for Video Quality

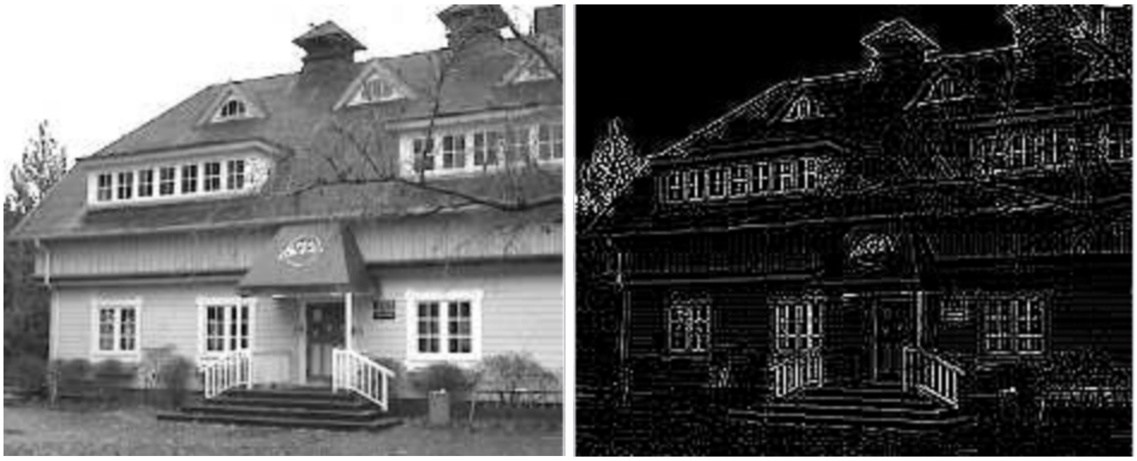
Given that the quality of C-SPAN videos change over time, in this section I include a control for video quality. However, before I introduce the additional control it is important to note that frame differencing naturally accounts for many of the issues related to video quality

Table S5: Are Democrats and Republicans Less Willing to Cross the Aisle After Party Votes? ("PHash = 100")

	Dependent variable:			
	Party Vote		Future Party Votes	
	Logistic Regression		*Tobit Regression*	
	(1)	(2)	(3)	(4)
Constant	0.247***	0.938***	0.036	0.012
	(0.048)	(0.181)	(0.022)	(0.052)
Structural Similarity	0.136**	0.240***	0.027	0.044**
	(0.057)	(0.070)	(0.017)	(0.019)
Passage Vote		0.285**		0.206***
		(0.141)		(0.039)
Amendment Vote		0.621***		0.267***
		(0.125)		(0.039)
Total Not Voting		−0.040***		−0.005**
		(0.007)		(0.002)
|Sponsor Ideology|		−1.095**		−0.385**
		(0.534)		(0.176)
Sponsor Seniority		−0.020		0.001
		(0.013)		(0.004)
Sponsor Party Leader		0.912**		0.025
		(0.384)		(0.097)
Election Year		−0.016		−0.051
		(0.117)		(0.047)
N	2,242	2,017	2,261	2,017
Log Lik	−1,532.814	−1,287.912	−1,748.501	−1,503.222
AIC	3,069.627	2,593.823	3,503.002	3,026.445

Note: These models are identical to the models reported in Table 1 except frames had to be sufficiently similar to at least 100 of the "good" frames to be used in the analysis. Dependent variable and the model used are reported above each column. Video motion is captured in the variable labeled "Structural Similarity." Positive values imply the frames within the video are more similar to one another, implying there is *less* motion. The unit of analysis is a given floor vote. The data has also been restricted to votes on House bills and resolutions. Standard errors clustered around the issue and day are reported in parentheses.

Figure S15: Using a Laplacian Filter for Edge Detection



(a) Original Image

(b) Original Image with Laplacian Filter

Note: This example was borrowed from Sinha (2010). Panel A includes the original image. Panel B shows the same image after a Laplacian filter has been applied.

since frames are being compared to one another. This essentially “cancels out” any problems associated with different frame borders, etc. which is why this measure was not included as a control in the main text.

With that said, in this subsection I include a control for video quality as an additional robustness check. Here, I measure video quality using the average Laplacian variation. Originally proposed by Pertuz, Puig and Garcia (2013), this measure first convolves a grayscale image using a 3 × 3 Laplacian kernel:

$$\begin{bmatrix} 10 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

When this is done, the second derivative of the image is returned, ultimately highlighting regions with greater pixel disparities (see Figure S15). This makes the Laplacian filter particularly useful for edge detection since edges often lie in regions where pixels dramatically change, such as from black (0) to white (1). Unsurprisingly, the variance of the second derivative will return higher values when there are several distinct edges and lower values when the edges are less defined. Videos with lower quality tend to have edges that are less defined, meaning they will have lower Laplacian variation.

In Table S6, I re-estimated the models found in Table 1 of the main text including the average Laplacian variation (see `Image Quality`). Regardless of the model, `Structural Similarity` is positive and statistically significant implying party votes are more likely to occur after videos in which *less* motion occurs. Collectively, these results provide strong

Table S6: When MC's Cross The Aisle, Future Party Votes Are Less Likely To Occur ("PHash = 0" with Video Quality)

	Dependent variable:					
	Future Party Votes					
	(1)			*(2)*		
Variable	$\hat{\beta}$	$\mathrm{SE}_{\hat{\beta}}$	95% CI	$\hat{\beta}$	$\mathrm{SE}_{\hat{\beta}}$	95% CI
Constant	0.199	0.055	$[0.092, 0.307]$	−0.654	0.151	$[-0.950, -0.358]$
Structural Similarity	0.073	0.025	$[0.025, 0.121]$	0.052	0.020	$[0.012, 0.092]$
Image Quality				0.000	0.000	$[0.000, 0.000]$
Previous Party Votes				0.483	0.093	$[0.302, 0.665]$
Passage Vote				0.004	0.048	$[-0.090, 0.099]$
Amendment Vote				0.478	0.060	$[0.360, 0.596]$
Total Not Voting				−0.020	0.003	$[-0.026, -0.014]$
|Sponsor Ideology|				0.279	0.163	$[-0.041, 0.600]$
Sponsor Seniority				−0.011	0.006	$[-0.022, 0.001]$
Sponsor Party Leader				−0.124	0.161	$[-0.440, 0.192]$
Election Year				1.157	0.097	$[0.966, 1.347]$
N	3,605			3,605		
Log Likelihood	−3,560.853			−3,172.448		
AIC	7,141.707			6,382.896		

Note: Models are identical to those reported in Table 1 except a control is included for video quality. Dependent variable is the number of party votes that occur after the current video divided by the total number of remaining votes. `Image Quality` described in Section S3.3 and `Structural Similarity` is described on pages S9–S12 in the SI. Positive values imply *less* video motion. Unit of analysis is a given floor vote. Since some House bills and resolutions have several votes, standard errors clustered around each piece of legislation (e.g., `HR` 820). 95% confidence intervals reported. All models estimated using the **tobit** function in **Stata** (v15).

evidence that my main results cannot be easily attributed to changes in the C-SPAN videos over time. Not only does frame differencing naturally address some of these concerns, but this subsection shows the results found in Table 1 remain the same when video quality is included as a control.

Although not the main purpose of this analysis including Laplacian variation as a control also serves as a robustness check for the number of MCs in the video. Since videos with more MCs tend to look more like a "blob" of people, then these videos tend to have less well defined edges as compared to videos where a single MC is standing on the floor. Given that, `Image Quality` also helps control for the number of MCs in the well. Indeed, in Table S6, Model 1 the coefficient associated with `Structural Similarity` increases 35.19 percent when `Image Quality` is included as a control which suggests this variable helps better isolate the relationship of interest.

S3.4 Majority Party Size

Another potentially confounding variable could be the size of the majority party. If the majority party is large, then party members are more likely to sit on opposite sides of the aisle simply because their party's side does not have enough seats. More recently, majorities have declined, so the decrease in cross-party discussions between floor votes could be due to the fact that some MCs no longer have to cross the aisle to speak with the opposition.

Table 1 in the main text includes fixed effects for each year. Since majority size has not repeated in the years of this study, then these fixed effects should effectively control for the size of the majority party. Table S7 provides a more explicit test. Here, the aforementioned fixed effects are replaced with the actual size of the majority party. Not only are the substantive results the same in these models, but the coefficient associated with `Structural Similarity` actually increases in Model 2.

S3.5 Polarized Legislative Speech

To demonstrate video motion is predictive of other forms of legislative behavior, I also created a measure of polarized legislative speech. If crossing the aisle captures the broader legislative environment, then I expect speeches to also be more polarized after floor votes in which Democrats and Republicans refuse to speak with one another. This argument is consistent with previous literature in which floor speeches have been used to capture party polarization (e.g., Jensen et al., 2012; Harris, 2005), suggesting such speeches are well-suited to check the robustness of the results reported in the main text.

In the models below, the dependent variable is derived from weighted versions of the "positive" and "negative" emotion categories from the Linguistic Inquiry Word Count (LIWC) dictionary. More specifically, for each MC, I only considered speeches which referred to the opposing party. I then determined the proportion of words which were positive and negative,

Table S7: When MC's Cross The Aisle, Future Party Votes Are Less Likely To Occur ("PHash = 0" with Majority Size)

		Dependent variable:				
		Future Party Votes				
		(1)			*(2)*	
Variable	$\hat{\beta}$	$SE_{\hat{\beta}}$	95% CI	$\hat{\beta}$	$SE_{\hat{\beta}}$	95% CI
Constant	0.138	0.074	$[-0.007, 0.284]$	-0.557	0.136	$[-0.824, -0.291]$
Structural Similarity	0.054	0.024	$[0.007, 0.101]$	0.063	0.021	$[0.022, 0.105]$
Majority Size				0.461	0.499	$[-0.517, 1.439]$
Previous Party Votes				0.507	0.093	$[0.324, 0.690]$
Passage Vote				0.005	0.048	$[-0.090, 0.100]$
Amendment Vote				0.499	0.059	$[0.382, 0.615]$
Total Not Voting				-0.021	0.003	$[-0.027, -0.015]$
|Sponsor Ideology|				0.455	0.150	$[0.160, 0.749]$
Sponsor Seniority				-0.008	0.006	$[-0.020, 0.003]$
Sponsor Party Leader				-0.124	0.158	$[-0.433, 0.185]$
Election Year				1.122	0.090	$[0.945, 1.299]$
N	3,605			3,605		
Log Likelihood	−3,621.416			−3,188.433		
AIC	7,250.832			6,400.866		

Note: Models are identical to those reported in Table 1 except a control is included for video quality. Dependent variable is the number of party votes that occur after the current video divided by the total number of remaining votes. `Image Quality` described in Section S3.3 and `Structural Similarity` is described on pages S9–S12 in the SI. Positive values imply *less* video motion. Unit of analysis is a given floor vote. Since some House bills and resolutions have several votes, standard errors clustered around each piece of legislation (e.g., `HR` 820). 95% confidence intervals reported. All models estimated using the **tobit** function in **Stata** (v15).

but instead of using the raw proportions, the proportions were weighted depending on how close each term was to opposing party references. This was done using the **wpct** function from the **weights** package in the **R** statistical software language. Using this proportion, I then determined whether speeches generally used more negative or positive words when referencing the opposing party.

Once these scores were created for each speech, I then determined the number of "polarized speeches" which occurred after the C-SPAN video of interest. All floor speech data was collected from a cached version of the *Capitol Words Project*. Speech times were determined using the *Congressional Record*. More specifically, I found the closest timestamps – in terms of words – from the speech. A weighted average was then created in which closer timestamps were weighted more heavily. Ultimately, this yielded start and stop times for 731,283 floor speeches between 1996–2014.

In the models below, `Future Party Speeches` was calculated using the number of speeches in which (1) the opposing party is referenced and (2) the weighted proportion of negative words is greater than the weighted proportion of positive words. Again, only speeches which occur after the current video are included. Similar to `Future Party Votes` in the main text, I also divided this count by the total number of remaining speeches. All control variables are described and justified in Table S1. Perhaps most importantly, a control is included for the number of party speeches that occurred *before* the current video divided by the total number of prior speeches (see `Previous Party Speeches`). Unit of analysis is the floor vote and only votes on House bills and resolutions are included.

The main results are reported in Table S8. Since `Future Party Speeches` ranges from 0 to 1, I report results from Tobit regressions with standard errors clustered around the bill under consideration. Beginning with Model 1, `Structural Similarity` is a positive and statistically significant at the 0.05-level, suggesting floor speeches become more polarized after videos of floor votes in which motion declines. Model 2 shows this result holds even when controlling for a number of factors, including the proportion of previous speeches in which the opposing party is referred to more negatively than positively.

To help interpret these results, predicted values were computed using the coefficients from Model 1. In the 112^{th} Congress, when `Structural Similarity` is allowed to very from $-\frac{1}{2}$ Standard Deviation (SD) (more motion) to $+\frac{1}{2}$ SD (less motion) the predicted number of party speeches increases 13.49 percent (0.059 to 0.067 speeches). Allowing the same variable to vary from -1 SD (more motion) to $+1$ SD (less motion) increases the predicted number of party speeches by 28.80 percent (0.056 to 0.072 speeches). Finally, in the 112^{th} Congress, when `Structural Similarity` is allowed to vary from -2 SD (more motion) to $+2$ SD (less motion) the predicted number of party speeches increases 65.90 percent (0.049 to 0.081 speeches), suggesting as video motion *decreases* party speeches are more likely to occur later that day.

Table S8: When MC's Cross The Aisle, Future Party Speeches Are Less Likely To Occur

		Dependent variable:				
		Future Party Speeches				
		(1)			*(2)*	
Variable	$\hat{\beta}$	SE$_{\hat{\beta}}$	95% CI	$\hat{\beta}$	SE$_{\hat{\beta}}$	95% CI
Constant	−0.978	0.187	$[-1.344, -0.612]$	−2.518	0.370	$[-3.244, -1.791]$
Structural Similarity	0.127	0.059	$[0.011, 0.243]$	0.129	0.051	$[0.028, 0.230]$
Previous Party Speeches				2.606	0.251	$[2.113, 3.099]$
Passage Vote				−0.165	0.105	$[-0.370, 0.041]$
Amendment Vote				0.211	0.124	$[-0.033, 0.455]$
Total Not Voting				−0.019	0.005	$[-0.029, -0.009]$
|Sponsor Ideology|				−0.237	0.381	$[-0.983, 0.509]$
Sponsor Seniority				−0.007	0.012	$[-0.030, 0.017]$
Sponsor Party Leader				−0.071	0.362	$[-0.781, 0.638]$
Election Year				2.831	0.268	$[2.306, 3.355]$
N	3,605			3,605		
Log Likelihood	−2,701.047			−2,285.044		
AIC	5,422.095			4,606.088		

Note: `Future Party Speeches` is the number of party speeches that occur after the current video divided by the total number of remaining speeches. This variable is described on page S32 in the SI. `Structural Similarity` is described on page 6 in the main text and pages S9–S12 in the SI. Positive values imply *less* video motion. Unit of analysis is a given floor vote. Since some House bills and resolutions have several votes, standard errors clustered around each piece of legislation (e.g., `HR` 820) are reported in parentheses. All models estimated using the **tobit** function in **Stata** (v15).

S3.6 Alternative Model Specifications

There are many ways to model the number of future party votes. In the main text, I use Tobit regressions to model the proportion of votes in which a majority of Democrats voted in the opposite direction as the majority of Republicans. In Table S9, the number of future party votes is modeled as a count. Here, the models are the same as those reported in Table 1, but the results are derived from negative binomial regressions. Controls were also added for the number of future (see `Future Votes`) and previous (see `Previous Votes`) votes. All other model choices are the same, including the standard errors which are still clustered around the bill under consideration.

Regardless of the model, `Structural Similarity` is positive and statistically significant which suggests the results reported in the main text cannot easily be attributed to modeling choices. In Table S10, I re-estimated the same models, but I also included a dummy variable indicating whether the vote was either the first or last on a given legislative day. This occurs periodically throughout the dataset. In the main text, these are recorded as zeros, but I wanted to estimate another version of the models in which these were given separate intercepts since the measure itself is not as clearly defined in these instances.

Similar to before, `Structural Similarity` is positive and statistically significant in these models which, again, suggests the results in the main text cannot be easily attributed to the way the models are specified. Undoubtedly, it is difficult to establish a causal relationship using observational data and this study does not claim to do so. However, given the evidence presented in the main text and SI, I think it is reasonable to suggest video motion is (somewhat) correlated with the degree to which Democrats and Republicans speak with one another after floor votes. I also show this behavior is predictive of future party votes and more polarized legislative speech which suggests crossing the aisle (or lack there of) has important downstream consequences.

S3.7 Party Leaders

On page 3 in the main text, two potential mechanisms are offered to explain why video motion should be predictive of future party votes. In the first, decreaseed socialization between the parties could signal to party members that the parties are at an impasse which would make bipartisan cooperation less likely that legislative day. In the second, conversations between Democrats and Republicans offer an opportunity for party members to work out their differences on upcoming votes.

Although it is difficult, if not impossible, to adjudicate between these two mechanisms, Table S11 gains some traction on the first mechanism by interacting `Sponsor Party Leader` with `Structural Similarity`. If decreased cross-party dialogue sends a partisan signal, then one would expect this signal to be stronger when the preceding bill was sponsored by a party leader. In Table S11 this interaction is not statistically significant which is evidence

Table S9: When MC's Cross The Aisle, Future Party Votes Are Less Likely To Occur

	Dependent variable:	
	Future Party Votes	
	(1)	(2)
Constant	−1.170	−0.807
	(0.070)	(0.171)
Structural Similarity	0.111	0.102
	(0.034)	(0.033)
Future Votes	0.216	0.203
	(0.013)	(0.012)
Previous Party Votes		0.108
		(0.041)
Previous Votes		−0.075
		(0.035)
Passage Vote		−0.010
		(0.067)
Amendment Vote		0.221
		(0.092)
Total Not Voting		−0.007
		(0.004)
|Sponsor Ideology|		−0.603
		(0.497)
Sponsor Seniority		−0.019
		(0.010)
Sponsor Party Leader		−0.005
		(0.243)
Election Year		−0.189
		(0.147)
Congress Fixed Effects	✓	✓
N	3,605	3,603
Log Likelihood	−4,100.092	−4,049.402
θ	1.808 (0.113)	1.916 (0.121)
AIC	8,206.184	8,122.804

Note: Dependent variable is the number of party votes that occur after the current video. `Structural Similarity` is described on page 6 in the main text and pages S9–S12 in the SI. Positive values imply *less* video motion. Unit of analysis is a given floor vote. Since some House bills and resolutions have several votes, standard errors clustered around each piece of legislation (e.g., `HR` 820) are reported in parentheses. All models estimated using the **Zelig** package in **R** (v3.6.0).

Table S10: When MC's Cross The Aisle, Future Party Votes Are Less Likely To Occur

	Dependent variable:	
	Future Party Votes	
	(1)	(2)
Constant	−1.170 (0.070)	−0.623 (0.185)
Structural Similarity	0.111 (0.034)	0.100 (0.033)
Future Votes	0.216 (0.013)	0.196 (0.012)
Time of Day		−0.012 (0.008)
Previous Party Votes		0.105 (0.041)
Previous Votes		−0.069 (0.035)
|Sponsor Ideology|		−0.579 (0.503)
Sponsor Seniority		−0.019 (0.010)
Sponsor Party Leader		−0.009 (0.243)
Passage Vote		−0.024 (0.067)
Amendment Vote		0.235 (0.093)
Total Not Voting		−0.007 (0.005)
Election Year		−0.173 (0.148)
Congress Fixed Effects	✓	✓
N	3,605	3,603
Log Likelihood	−4,100.092	−4,047.434
θ	1.808 (0.113)	1.938 (0.124)
AIC	8,206.184	8,120.867

Note: Dependent variable is the number of party votes that occur after the current video. `Structural Similarity` is described on page 6 in the main text and pages S9–S12 in the SI. Positive values imply *less* video motion. Unit of analysis is a given floor vote. Since some House bills and resolutions have several votes, standard errors clustered around each piece of legislation (e.g., `HR` 820) are reported in parentheses. All models estimated using the **Zelig** package in **R** (v3.6.0).

against the first mechanism, but this is far from a definitive test and should be taken with the appropriate level of skepticism.

Table S11: No Significant Interaction Between Video Motion and Whether The Bill Was Sponsored By A Party Leader

	Dependent variable:					
	Future Party Votes					
	(1)			*(2)*		
Variable	$\hat{\beta}$	$\text{SE}_{\hat{\beta}}$	95% CI	$\hat{\beta}$	$\text{SE}_{\hat{\beta}}$	95% CI
Constant	0.204	0.055	$[0.096, 0.311]$	-0.521	0.143	$[-0.802, -0.240]$
Structural Similarity	0.071	0.025	$[0.023, 0.120]$	0.053	0.021	$[0.013, 0.093]$
Sponsor Party Leader	-0.255	0.171	$[-0.590, 0.081]$	-0.139	0.153	$[-0.438, 0.160]$
Previous Party Votes				0.492	0.093	$[0.309, 0.675]$
Passage Vote				0.008	0.048	$[-0.087, 0.103]$
Amendment Vote				0.498	0.060	$[0.371, 0.607]$
Total Not Voting				-0.020	0.003	$[-0.026, -0.015]$
|Sponsor Ideology|				0.297	0.164	$[-0.025, 0.619]$
Sponsor Seniority				-0.009	0.006	$[-0.021, 0.002]$
Election Year				1.158	0.097	$[0.967, 1.349]$
Structural Similarity × Sponsor Party Leader	0.110	0.145	$[-0.174, 0.395]$	0.067	0.127	$[-0.181, 0.316]$
N	3,605			3,605		
Log Likelihood	−3,559.098			−3,178.744		
AIC	7,142.195			6,395.487		

Note: Dependent variable is the number of party votes that occur after the current video divided by the total number of remaining votes. `Structural Similarity` is described on page 9 in the main text. Positive values imply *less* video motion. Unit of analysis is a given floor vote. Since some House bills and resolutions have several votes, standard errors are clustered around each piece of legislation (e.g., `HR` 820). 95% confidence intervals are also reported. All models include Congress fixed-effects and were estimated using the **tobit** function in **Stata** (v15).

S3.8 Endogeneity Checks

Although I do not aim to say definitively how the "social fabric" of Congress influences legislative behavior, some may be interested in the causal ordering. Does the lack of inter-party dialogue lead to party-line voting? Or does party-line voting create bad blood between the political parties which precludes inter-party dialogue between floor votes? Both could also be influenced by polarization in the electorate which leads to party-line voting in Congress while also making opposing MCs less likely to speak to each other. Each of these alternative explanations are considered in the tables below.

Table S12: Current Video Motion Does Not Predict Previous Party Votes (Endogeneity Check #1)

	Dependent variable:					
	Previous Party Votes					
	(1)			*(2)*		
Variable	$\hat{\beta}$	$\text{SE}_{\hat{\beta}}$	95% CI	$\hat{\beta}$	$\text{SE}_{\hat{\beta}}$	95% CI
Constant	-0.059	0.059	$[-0.175, 0.057]$	0.518	0.076	$[0.369, 0.668]$
Structural Similarity	−0.016	0.014	$[-0.043, 0.012]$	0.016	0.010	$[-0.004, 0.036]$
Future Party Votes				0.201	0.038	$[0.126, 0.277]$
Passage Vote				0.273	0.027	$[0.219, 0.326]$
Amendment Vote				0.189	0.034	$[0.123, 0.255]$
Total Not Voting				−0.011	0.001	$[-0.014, -0.008]$
|Sponsor Ideology|				0.133	0.117	$[-0.096, 0.363]$
Sponsor Seniority				0.002	0.003	$[-0.005, 0.008]$
Sponsor Party Leader				0.337	0.082	$[0.175, 0.499]$
Election Year				−3.747	0.131	$[-4.004, -3.489]$
N	3,605			3,605		
Log Likelihood	−3,387.449			−2,258.445		
AIC	6,794.898			4,552.890		

Note: Dependent variable is the number of party votes that occur after the current video divided by the total number of remaining votes. `Structural Similarity` is described on page 9 in the main text. Positive values imply *less* video motion. Unit of analysis is a given floor vote. Since some House bills and resolutions have several votes, standard errors are clustered around each piece of legislation (e.g., `HR` 820). 95% confidence intervals are also reported. All models include Congress fixed-effects and were estimated using the **tobit** function in **Stata** (v15).

In Table S12 the dependent variable is the percent of previous votes that were party votes. If the relationship between `Future Party Votes` and `Structural Similarity` is due to a general trend of party voting happing on the same legislative day, then `Structural Similarity` should also predict `Previous Party Votes`, but this is not the case. Not only is the relationship far from statistically significant, but the sign of the coefficient changes from one model to the next which gives additional evidence that there is no meaningful relationship between the video motion at time t and the level of party voting at time $t-1$,

$t-2, \ldots t-n$ which is consistent with the original interpretation of the results provided in the main text.

Table S13: Previous Party Votes Do Not Predict Current Video Motion (Endogeneity Check #2)

	Dependent variable:					
	Structural Similarity					
	(1)			*(2)*		
Variable	$\hat{\beta}$	$\text{SE}_{\hat{\beta}}$	95% CI	$\hat{\beta}$	$\text{SE}_{\hat{\beta}}$	95% CI
Constant	-0.245	0.034	$[-0.312, -0.179]$	-0.423	0.077	$[-0.574, -0.272]$
Previous Party Votes	-0.104	0.041	$[-0.185, -0.022]$	0.058	0.052	$[-0.043, 0.160]$
Passage Vote				-0.143	0.037	$[-0.216, -0.070]$
Amendment Vote				-0.046	0.037	$[-0.118, 0.027]$
Total Not Voting				0.004	0.001	$[0.001, 0.006]$
|Sponsor Ideology|				0.139	0.096	$[-0.050, 0.328]$
Sponsor Seniority				-0.004	0.003	$[-0.011, 0.002]$
Sponsor Party Leader				0.011	0.087	$[-0.160, 0.181]$
Election Year				0.197	0.055	$[0.089, 0.305]$
N	3,605			3,605		
Log Likelihood	−3,108.769			−2,258.445		
AIC	6,237.538			4,552.890		

Note: Dependent variable is the number of party votes that occur after the current video divided by the total number of remaining votes. `Structural Similarity` is described on page 9 in the main text. Positive values imply *less* video motion. Unit of analysis is a given floor vote. Since some House bills and resolutions have several votes, standard errors are clustered around each piece of legislation (e.g., `HR` 820). 95% confidence intervals are also reported. All models include Congress fixed-effects and were estimated using the **tobit** function in **Stata** (v15).

These results are further supported in Table S13 which shows that `Previous Party Votes` is not a statistically significant predictor of `Structural Similarity`. When these results are compared to those reported in Table 1 and Tables S1-S10, more evidence suggests `Structural Similarity` is predictive of party voting and not the other way around. However, these results should be taken with some degree of skepticism since these are by no means definitive causal tests.

As mentioned above, the results outlined throughout the main text and SI could be attributed to polarization in the electorate. If this were the case, then `Previous Party Votes`, `Future Party Votes`, and `Structural Similarity` should consistently predict one another, but this is not what was found in Tables S12 and S13. Table S14 provides some

Table S14: Number of Independents Does Not Explain The Results (Endogeneity Check #3)

			Dependent variable:			
			Future Party Votes			
		(Lag 1)			*(Lag 2)*	
Variable	$\hat{\beta}$	$\text{SE}_{\hat{\beta}}$	95% CI	$\hat{\beta}$	$\text{SE}_{\hat{\beta}}$	95% CI
Constant	−0.454	0.574	$[-1.580, 0.672]$	−0.374	0.558	$[-1.468, 0.720]$
Structural Similarity	0.068	0.021	$[0.026, 0.110]$	0.068	0.021	$[0.026, 0.110]$
Percent Independent	−0.003	0.018	$[-0.037, 0.032]$	−0.005	0.017	$[-0.039, 0.029]$
Previous Party Votes	0.523	0.091	$[0.344, 0.703]$	0.523	0.092	$[0.343, 0.703]$
Passage Vote	0.003	0.048	$[-0.092, 0.098]$	0.003	0.048	$[-0.091, 0.098]$
Amendment Vote	0.497	0.059	$[0.381, 0.614]$	0.498	0.060	$[0.381, 0.615]$
Total Not Voting	−0.021	0.003	$[-0.027, -0.015]$	−0.021	0.003	$[-0.027, -0.015]$
|Sponsor Ideology|	0.484	0.149	$[0.192, 0.776]$	0.486	0.149	$[0.194, 0.777]$
Sponsor Seniority	−0.007	0.006	$[-0.019, 0.004]$	−0.008	0.006	$[-0.019, 0.003]$
Sponsor Party Leader	−0.126	0.158	$[-0.436, 0.183]$	−0.127	0.158	$[-0.438, 0.183]$
Election Year	1.116	0.092	$[0.935, 1.297]$	1.118	0.092	$[0.938, 1.298]$
N	3,605			3,605		
Log Likelihood	−3,189.198			−3,189.198		
AIC	6,402.396			6,402.396		

Note: Dependent variable is the number of party votes that occur after the current video divided by the total number of remaining votes. `Structural Similarity` is described on page 9 in the main text. Positive values imply *less* video motion. Unit of analysis is a given floor vote. Since some House bills and resolutions have several votes, standard errors are clustered around each piece of legislation (e.g., `HR` 820). 95% confidence intervals are also reported. All models include Congress fixed-effects and were estimated using the **tobit** function in **Stata** (v15).

additional support by including the percent of Americans who identified as Independents according to Gallup[S9] as an additional control. Even with this variable – labeled `Percent Independent` – is included as a control `Structural Similarity` is still a statistically significant predictor of `Future Party Votes`.

References

Fowler, James H. 2006. "Connecting the Congress: A Study of Cosponsorship Networks." *Political Analysis* 14(4):456–487.

Harris, Douglas B. 2005. "Orchestrating Party Talk: A Party-Based View of One-Minute Speeches in the House of Representatives." *Legislative Studies Quarterly* 30(1):127–141.

Hu, Weiming, Tieniu Tan, Liang Wang and Steve Maybank. 2004. "A Survey on Visual Surveillance of Object Motion and Behaviors." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 34(3):334–352.

Jensen, Jacob, Suresh Naidu, Ethan Kaplan, Laurence Wilse-Samson, David Gergen, Michael Zuckerman and Arthur Spirling. 2012. "Political polarization and the dynamics of political language: Evidence from 130 years of partisan speech [with comments and discussion]." *Brookings Papers on Economic Activity* pp. 1–81.

Pertuz, Said, Domenec Puig and Miguel Angel Garcia. 2013. "Analysis of Focus Measure Operators for Shape-from-Focus." *Pattern Recognition* 46(5):1415–1432.

Seshadrinathan, Kalpana and Alan C Bovik. 2007. A structural similarity metric for video based on motion models. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*. Vol. 1 IEEE pp. I–869.

Wang, Zhou and Alan C. Bovik. 2009. "Mean Squared Error: Love it or Leave it? A New Look at Signal Fidelity Measures." *IEEE Signal Processing Magazine* 26(1):98–117.

Wang, Zhou, Alan C. Bovik, Hamid R. Sheikh and Eero P. Simoncelli. 2004. "Image Quality Assessment: From Error Visibility to Structural Similarity." *IEEE Transactions on Image Processing* 13(4):600–612.

Yilmaz, Alper, Omar Javed and Mubarak Shah. 2006. "Object Tracking: A Survey." *Acm computing surveys (CSUR)* 38(4):13.

[S9] `https://news.gallup.com/poll/15370/party-affiliation.aspx`

Zauner, Christoph. 2010. Implementation and Benchmarking of Perceptual Image Hash Functions. Master's thesis Upper Austria University of Applied Sciences Hagenberg: .

Zhou, Suiping, Dan Chen, Wentong Cai, Linbo Luo, Malcolm Yoke, Feng Tian, Victor Su-Han Tay, Darren Wee Sze Ong and Benjamin D. Hamilton. 2010. "Crowd Modeling and Simulation Techniques." *ACM Transactions on Modeling and Computer Simulation* 20(4):2001–2035.