

Using Motion Detection to Measure Social Polarization in the U.S. House of Representatives

Bryce J. Dietrich* 

Department of Political Science, University of Iowa, 341 Schaeffer Hall, Iowa City, IA 52242, USA.
Email: bryce-dietrich@uiowa.edu

Abstract

Although previous scholars have used image data to answer important political science questions, less attention has been paid to video-based measures. In this study, I use motion detection to understand the extent to which members of Congress (MCs) literally cross the aisle, but motion detection can be used to study a wide range of political phenomena, like protests, political speeches, campaign events, or oral arguments. I find not only are Democrats and Republicans less willing to literally cross the aisle, but this behavior is also predictive of future party voting, even when previous party voting is included as a control. However, this is one of the many ways motion detection can be used by social scientists. In this way, the present study is not the end, but the beginning of an important new line of research in which video data is more actively used in social science research.

Keywords: video analysis, Congress, polarization

1 Introduction

If an image is worth a thousand words, what is a video worth? Although previous scholars have used image data (e.g., Torres 2018; Casas and Williams 2019; Xi *et al.* 2019) to answer important political science questions, less attention has been paid to video-based measures. Using the largest collection of Cable-Satellite Public Affairs Network (C-SPAN) videos ever compiled, this paper introduces motion detection as a way to begin to use video-as-data. In this paper, I use this measure to understand the extent to which members of Congress (MCs) literally cross the aisle, but it could be used to study a wide range of political phenomena, like nonverbal displays during political speech (Koppensteiner and Grammer 2010) or the intensity of police–citizen encounters (Makin *et al.* 2019). In this way, the present study is not the end, but the beginning of an important new line of research in which videos are viewed as more than collections of images.

Motion detection is a useful starting place since it underlies other computer vision techniques, like gait analysis (Aggarwal and Cai 1999), object (Prajapati and Galiyawala 2015) and pedestrian tracking (Li *et al.* 2014), or scene detection (Koprinska and Carrato 2001). Although there are a variety of ways to detect motion (Manchanda and Sharma 2016), this study uses frame differencing because it is less computationally intensive and can be applied to low resolution videos like many of the videos found on C-SPAN's website or video sharing sites like YouTube. For example, optical flow techniques have also been used extensively in the literature (Shafie, Hafiz, and Ali 2009), but these methods typically require graphical processing units (GPUs) that are unavailable to many social scientists which is why frame differencing is offered as a useful way to begin video-based research.

Although scholars in other fields have used frame differencing to understand interpersonal dynamics (e.g., Paxton and Dale 2013; Ramseyer and Tschacher 2014), this technique has not yet been used in political science. In this paper, frame differencing is used to measure the extent to which MCs speak with members of the opposing party after floor votes. Not only is this a behavior

Political Analysis (2021)
vol. 29: 250–259
DOI: 10.1017/pan.2020.25

Published
11 November 2020

Corresponding author
Bryce J. Dietrich

Edited by
Jeff Gill

© The Author(s) 2020. Published
by Cambridge University Press
on behalf of the Society for
Political Methodology.

*The keywords have been corrected. A corrigendum notice detailing this change was also published (DOI: [10.1017/pan.2021.8](https://doi.org/10.1017/pan.2021.8)).

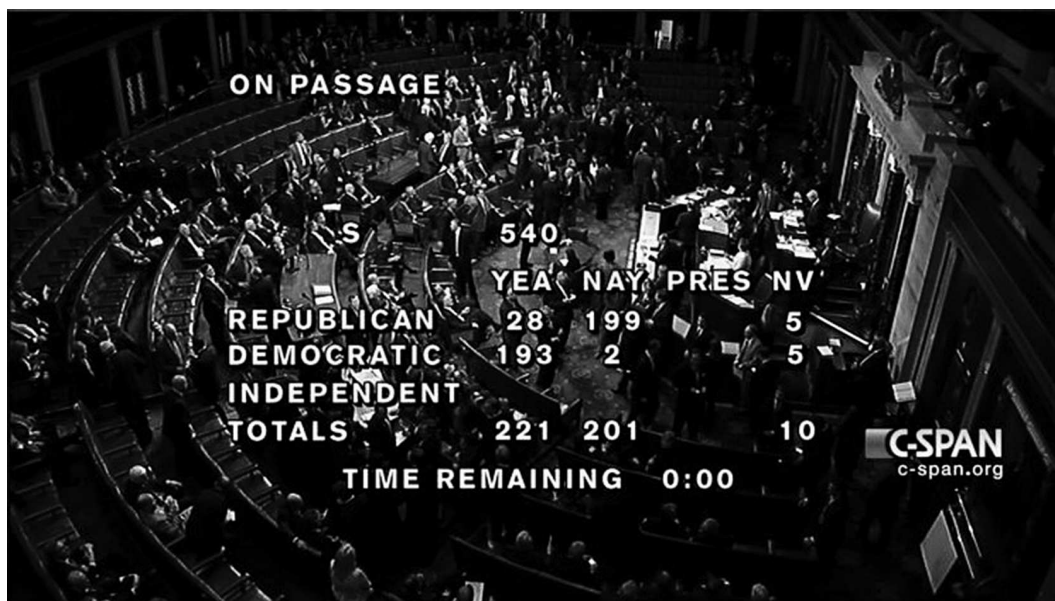


Figure 1. Overhead shot of members of Congress mingling after a roll-call vote. Not only does this shot show all of the social interactions that take place after a floor vote, but it is a quintessential part of C-SPAN coverage. All the analyses presented below consider videos similar to the frame shown here.

of interest to a number of legislative scholars (Masket 2008; Cohen and Malloy 2014), but also as I will show below, my video-based measure yields reasonable results which suggests it could be a useful alternative to the more time- and labor-intensive approaches used by previous scholars (e.g., Caldeira and Patterson 1987). Ultimately, I find Democrats and Republicans are increasingly less willing to talk to one another after floor votes and this behavior is predictive of future party votes, but this is one of the many ways motion detection can be used by social scientists.

2 The Congressional Workplace

Congress is a unique workplace in that (a) interacting with people with different political views is often necessary to achieve personal and institutional objectives, yet (b) people are physically sorted on the basis of their political views. One of the few formal settings where Democrats and Republicans informally talk to one another (Hughes *et al.* 1976) and “mill around” is the well of the U.S. House of Representatives (Green and Hogan 1982). These “water cooler conversations” have been found to have a moderating effect in other contexts (Mutz and Mondak 2006), but the mingling shown in Figure 1 has never been studied in political science, despite previous scholars emphasizing the importance of spatial proximity in legislative settings (Masket 2008; Cohen and Malloy 2014).

Kirkland (2011)’s distinction between strong and weak legislative ties helps explain why such informal conversations may be important to legislative outcomes. When ties are strong, new information is rarely exchanged, preventing the size of the legislative coalition from expanding. Since weak legislative ties originate from less frequent interactions, when those interactions do take place, information is more likely to be novel and ultimately shared which helps lay the foundation for future cooperation. Although previous scholars have demonstrated the importance of social interactions in other legislative settings (e.g., Caldeira and Patterson 1987), this application is the first to suggest such relationships may also be cultivated on the House floor.

In Section S2.3 of the SI, an agent-based model is used to show that video motion increases when people walk from one side of the room to the other in order to talk.¹ “At present Democrats

¹ The videos associated with these simulations are also helpful in understanding what “video motion” looks like. Please compare the video of the partisan (<https://youtu.be/kNBmFpqdBkw>) and bipartisan simulation

and Republicans literally sit ‘across the aisle’ from one another,” with Democrats sitting on the east side of the Chamber to the right of the Speaker of the House and Republicans sitting on the Speaker’s left (Gibson 2010, 21). Thus, MCs have to walk a greater distance to speak with members of the opposition, meaning videos should contain more motion when these interactions occur.² This leads to the following expectation: *on average, when the videos of social interactions immediately after floor votes display less motion, party votes are more likely to occur later that day.*

I offer two potential mechanisms, but do not make any strong causal claims. First, the socialization that occurs between votes (or lack thereof) could serve as an important signal to the rank-and-file. To use Kirkland (2011)’s example, just as Ted Kennedy (D-MA) working with Orrin Hatch (R-UT) in the 108th Senate could help convince some Republicans to take a closer look at Kennedy’s legislation, when Democrats and Republicans refuse to speak with one another after a floor vote there is even less incentive for MCs to work with one another which could increase party-line voting later that day.

Second, the conversations that occur between votes can be thought of as an example of “weak ties,” in that they are novel, timely, and provide information sharing opportunities. When these conversations occur *between* Democrats and Republicans they offer an opportunity for both sides to work out potential differences on later votes. Conversely, when such conversations are polarized—meaning they happen on opposite sides of the room—then there are comparatively fewer opportunities to find common ground, ultimately producing more party votes.

However, this research letter does not aim to resolve a debate that has existed in some fashion since perhaps Matthews (1959). Rather than saying definitively how the “social fabric” of Congress influences legislative behavior (Cho and Fowler 2010, 125), the degree to which Democrats and Republicans literally cross the aisle is offered as one way motion detection can be used for social science research. Scholars should use this measure to answer their own research questions and this application is meant to do nothing more than motivate this future work.

3 Data and Measure

3.1 C-SPAN Videos

This study employs the largest collection of video data—6,526 videos or 1,413 hours of C-SPAN coverage—ever used in political science research. 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. Segments similar to the frame shown in Figure 1 are the focus of my analysis. Not only are these overhead shots quintessential examples of C-SPAN coverage, but they also show *all* of the interactions that occur after floor votes while other camera angles only show a small fraction of these discussions.³

Figure 2 shows how these shots were identified in the 6,411,694 frames 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.⁴ A review of the output and an initial validation exercise shows my approach yields reasonable results.⁵

(<https://youtu.be/Ot1xerXV9qw>). The latter video has significantly more motion (see Figure S12) because the MCs have to walk a further distance.

2 In Section S2.4 of the SI, the debate surrounding the 2002 Bipartisan Campaign Reform Act is also used to show this general relationship.

3 See pages S2–S3 in the SI for more details.

4 All models were also re-estimated varying the number of “good” frames. Those results are reported on pages S24–S29 in the SI.

5 See pages S6–S9 in the SI for more details.

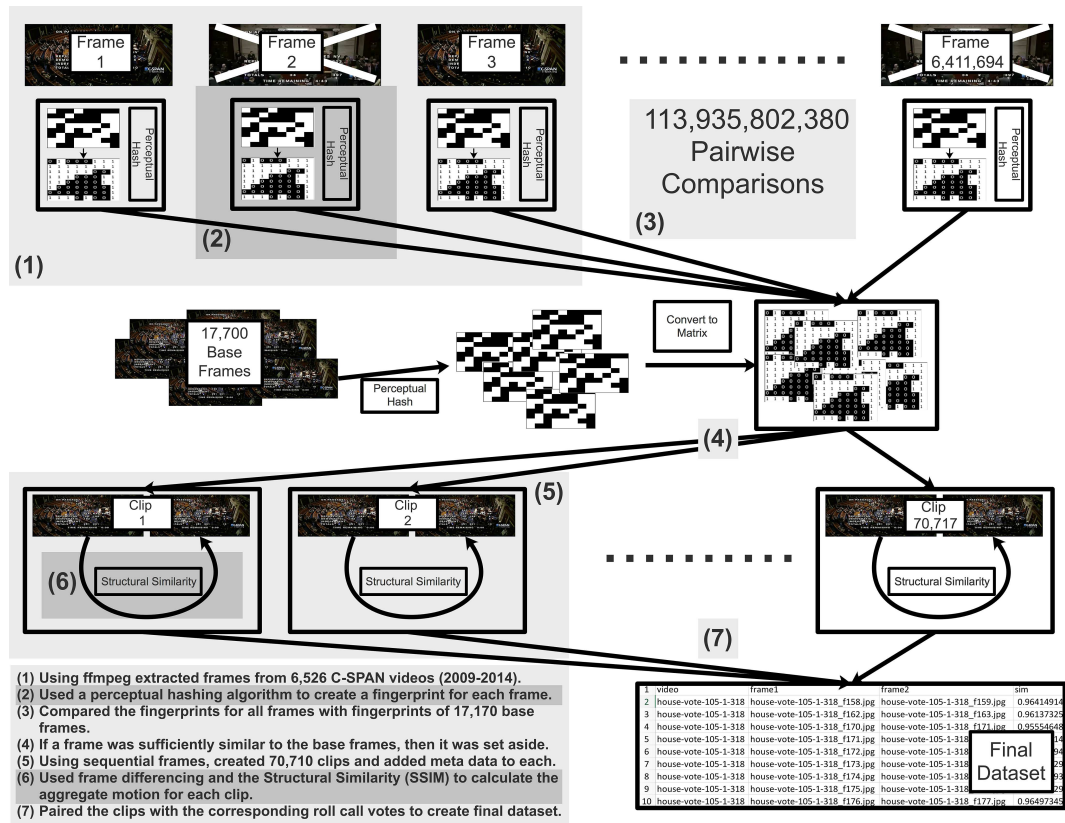


Figure 2. Figure explaining motion detection technique and how the overhead shots were extracted from the C-SPAN videos. Please see Section S1 in the Supplemental Information for more details about how the overhead shots were extracted and video motion was detected.

3.2 Motion Detection

In total, 70,717 clips were created with each video containing around 17 relevant clips, meaning within each video there were approximately 17 uninterrupted sequences of frames like Figure 1. Similar to Seshadrinathan and Bovik (2007), frames were differenced using the structural similarity (SSIM) between each frame. As explained in Section S1.4 of the SI, this measure—developed by Wang *et al.* (2004)—has many desirable properties, which is why it was used for this study. Ultimately, the overall motion of the video is the average SSIM across all sequential pairs of frames, which was calculated for the 2,935 C-SPAN videos in which at least one clip with two or more overhead shots could be identified.

Generally speaking, if \mathbf{x} and \mathbf{y} are the pixel matrices associated with two images, then the SSIM can be defined as:

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

where μ_x and μ_y are the means of \mathbf{x} and \mathbf{y} , respectively. Similarly, σ_x and σ_y are the SDs of \mathbf{x} and \mathbf{y} , leaving σ_{xy} as the cross correlation of \mathbf{x} and \mathbf{y} . C_1 and C_2 are small stabilizing constants and are defined in Python as follows:

$$C_1 = (0.01 \times (\max_{\mathbf{x}, \mathbf{y}} - \min_{\mathbf{x}, \mathbf{y}}))^2$$

$$C_2 = (0.03 \times (\max_{\mathbf{x}, \mathbf{y}} - \min_{\mathbf{x}, \mathbf{y}}))^2,$$

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

3.3 Validation

To help interpret the measure, I scaled the average SSIM to SDs above and below the mean—with positive values implying *less* motion.⁶ In order to validate the measure, 500 clips were randomly sampled. The extracted frames were then coded for whether someone was (1) or was not (0) walking in the well of the House. Figure S8 in the SI reports the results. Regardless of the Congress, frames with MCs walking in the well of the House are *less* similar, implying there is significantly *more* motion ($p < 0.001$).⁷

Next, 100 clips were randomly sampled from those used above and an undergraduate research assistant manually tracked each MC using the Fiji distribution of *ImageJ*.⁸ Abstract representations of each clip were then created using the tracking information. Next, I sequentially removed each MC and created new videos. The motion of these videos were then compared, allowing me to determine the degree to which each MC's walking patterns influenced the original video's overall motion. Again, MCs were coded for whether they did (1) or did not (0) cross the aisle. Figure S10 in the SI shows MCs produce significantly ($p < 0.001$) *more* motion when crossing the aisle, further validating my measure.⁹

Finally, I plotted average video motion against the absolute difference between the mean Democratic and Republican DW-NOMINATE scores for the 105th–112th Congresses. The results are shown in Figure 3. Ultimately, as time progresses (1) the absolute difference between Democratic and Republican DW-NOMINATE scores increases and (2) the videos of MCs mingling after floor votes have *more* SSIM implying *less* motion. Not only are these two variables highly correlated ($\rho = 0.79$), but Pearson's correlation coefficient is also statistically significant at the .02-level ($t = 3.18$, $df = 6$, and $p < .02$). In the SI, this measure is also validated using simulated bipartisan social interactions (see Section S2.3) and by predicting polarized legislative speech (see Section S3.5).

This latter result helps underline why video motion is a uniquely important measure. For example, Table S8 in the SI shows the ideology of the bill sponsor—as measured by DW-NOMINATE—is not predictive of polarized legislative speech. This is not to say that ideology plays no role in legislative speech, but rather to suggest that instead of being a replacement to these more traditional measures, video motion actually provides an important compliment. Indeed, interpersonal interactions are one of the many facets of legislative life (Patterson 1959; Bogue and Marlaire 1975; Caldeira and Patterson 1987). What I provide is a way to objectively measure such behavior using C-SPAN video data which is less time- and labor-intensive as compared to previous approaches.

3.4 Modeling Strategy

In the models below, my independent variable—called *Structural Similarity*—is the standardized SSIM described above. The dependent variable—called *Future Party Votes*—is the number of party votes that occur *after* the current video divided by the total number of remaining votes. All control variables are described and justified in Table S1 of the SI. Perhaps most importantly, a control is included for the number of party votes that occurred *before* the current video divided by the total number of prior votes.¹⁰ In Table 1, this variable is labeled *Previous Party*

⁶ See pages S9–S12 in the SI for more details.

⁷ See pages S13–S15 in the SI for more details.

⁸ <https://fiji.sc/>

⁹ See pages S15–S18 in the SI for more details.

¹⁰ Another important variable is *Not Voting* which is a continuous variable indicating the number of MCs who did not vote. Not only does this variable capture the relative importance of a vote, but I also found it was correlated with the number

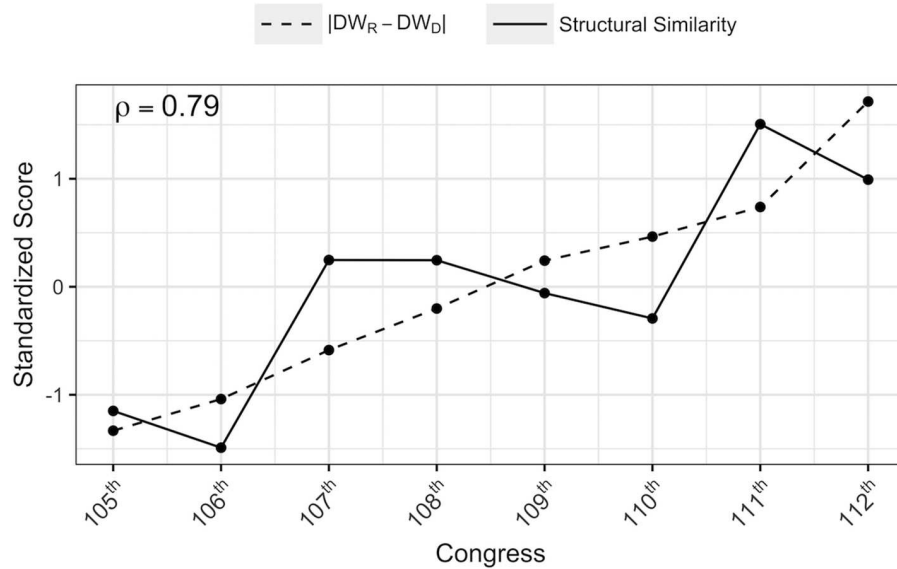


Figure 3. Plot of the average structural similarity for the 105th–112th Congress. Positive values imply the frames are more similar to one another implying there is *less* motion. The correlation between video motion and the absolute difference in Democratic and Republican DW-NOMINATE scores is reported in the top left. Both variables were standardized to standard deviations above/below the mean

Votes, the unit of analysis is the floor vote, and only votes on House bills and resolutions are included. Finally, all estimates are from Tobit regressions with standard errors clustered around the bill/resolution since the dependent variable ranges from 0 to 1 and the same bill/resolution can appear multiple times.

4 Results

Beginning with Model 1, *Structural Similarity* is positive and statistically significant ($p < 0.001$), suggesting party votes are *more* likely after videos of floor votes in which motion *declines*. Model 2 shows this result holds even when controlling for a number of factors, including *Previous Party Votes*. Here, when additional controls are included the positive coefficient associated with *Structural Similarity* (0.054) is close to three times the standard error (0.020), meaning as video motion *declines*—implied by the greater similarity between frames—party votes are significantly *more* likely to occur. These results are replicated a number of times in the SI using a variety of model specifications and various robustness checks which all lead to the same substantive conclusion.¹¹

To help interpret these results, predicted values were computed using the coefficients from Model 1. In the 112th Congress, when *Structural Similarity* is allowed to vary from $-\frac{1}{2}$ standard deviation (SD) (more motion) to $+\frac{1}{2}$ SD (less motion) the predicted number of party votes increases 7.31 percentage points (0.34 to 0.41 votes). Allowing the same variable to vary from -1 SD (more motion) to $+1$ SD (less motion) increases the predicted number of party votes by 14.61 percentage points (0.30 to 0.45 votes). Finally, in the 112th Congress, when *Structural*

of MCs on the screen. *Passage Vote* was included as a control for similar reasons since these votes tend to be more important and the videos tend to have more MCs present. Finally, on page S30 in the SI, I also explain why including Laplacian variation—a variable related to “image quality”—as a control also helps address any concerns about the number of MCs on the floor.

¹¹ The results are replicated in Section S3.1 excluding extreme observations. In Sections S3.2 and S3.3, I also show neither the clip configuration nor systematic changes in image quality can account for my main results. As explained in Section S3.4, the size of the majority party could also influence the degree to which Democrats and Republicans cross the aisle. In the main text, I already include a partial control in the form of congress fixed effects, but a more definitive test is provided in the SI. In these models, my main results still hold even when a control is included for the relative size of the majority party. Finally, Section S3.6 shows my results hold when all coefficients are estimated using negative binomial regressions.

Table 1. When MC's cross the aisle, future party votes are less likely to occur.

Variable	Dependent variable					
	Future party votes					
	(1)			(2)		
	$\hat{\beta}$	SE $\hat{\beta}$	95% CI	$\hat{\beta}$	SE $\hat{\beta}$	95% CI
Constant	0.199	0.055	[0.092, 0.307]	-0.521	0.143	[-0.802, -0.240]
Structural similarity	0.073	0.025	[0.025, 0.121]	0.054	0.020	[0.014, 0.094]
Previous party votes				0.492	0.093	[0.310, 0.675]
Passage vote				0.008	0.048	[-0.087, 0.103]
Amendment vote				0.489	0.060	[0.371, 0.607]
Total not voting				-0.020	0.003	[-0.026, -0.015]
Sponsor ideology				0.296	0.164	[-0.026, 0.617]
Sponsor seniority				-0.009	0.006	[-0.021, 0.002]
Sponsor party leader				-0.123	0.159	[-0.436, 0.189]
Election year				1.159	0.097	[0.968, 1.349]
N	3,605			3,605		
Log likelihood	-3,560.853			-3,178.87		
AIC	7,141.707			6,393.740		

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 19. 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). See Dietrich (2020) for data and replication code.

Similarity is allowed to vary from -2 SD (more motion) to +2 SD (less motion) the predicted number of party votes increases 29.22 percentage points (0.23 to 0.52 votes), suggesting as video motion decreases party votes are more likely to occur later that day.

Although this application is meant to demonstrate one way frame differencing could be used for social science research, many 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. It is undoubtedly difficult, if not impossible, to answer such questions using the observational design employed in this study, nor is that the goal of this research letter. However, in the SI, several additional tests were conducted to try to provide some initial answers, even though additional work must be done.

First, Table S12 shows that Structural Similarity is *not* a statistically significant predictor of Previous Party Votes, meaning the motion of a video at time t does not predict party votes at time $t-1, t-2, \dots, t-n$. Similarly, Table S13 shows Previous Party Votes is *not* a statistically significant predictor of Structural Similarity. If the relationship between party-line voting and inter-party discussions is due to polarization in the electorate, then Previous Party Votes, Future Party Votes, and Structural Similarity should consistently predict one another. Not only is this not what I found, but Table S14 shows my main results also hold when an additional control is included for the percent of voters who identify as Independent. Collectively, 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. Finally, Table S11 shows there is no significant interaction between Structural Similarity and Sponsor Party Leader which is inconsistent with the partisan signaling mechanism outlined on

page 8, but these tests are only suggestive and should be taken with the appropriate amount of skepticism.

5 Discussion and Conclusion

Although previous scholars have studied image data (e.g., Torres 2018; Casas and Williams 2019; Xi *et al.* 2019), the methodology I introduce is unique to video data, which includes television broadcasts, social media clips, and recorded deliberations, among other things. Rather than thinking of these videos as collections of images, this study offers motion detection as one of the many ways the temporal dynamics of videos can be used to answer important social science questions. In this way, this research letter should be thought of as an important first step towards better harnessing the use of video-as-data.

Frame differencing is not as computationally intensive and can be successfully used with both low- and high-resolution videos which is why it is offered as a useful starting place for this line of inquiry. Although my application is compelling, it is only used to highlight the methodology which can be used by a wide range of scholars. For example, the “energy” of campaign events or party conventions could be quantified by the degree to which videos change from one frame to the next (Zhong *et al.* 2007). Changes in motion could also be used to detect crowd panic (de Almeida *et al.* 2016), like what often occurs when protesters and counter-protesters clash.

More localized motion has also been shown to be associated with nonverbal displays (Murthy and Jadon 2009) which would be of interest to scholars of elite speech, especially those who emphasize the importance of the *way* words are spoken in the U.S. House of Representatives (Dietrich, Hayes, and O’Brien 2019) and on the Supreme Court (Dietrich *et al.* 2019). Finally, emotional contagion can also be measured using frame differencing (Paxton and Dale 2013; Ramseyer and Tschacher 2014). This relationship would be of interest to various political psychologists who have studied emotion using both experimental (Huddy, Mason, and Aarøe 2015) and observational designs (Valentino *et al.* 2011).

My results are also substantively important since they suggest the social environment on Capitol Hill (or lack thereof) may be more important than previously thought. Although these results are far from definitive, the fact they are suggestive lays an important foundation for future research. The approaches used by previous scholars to understand the “social fabric” of Congress are more labor- and time-intensive than the methodology introduced in this paper (e.g., Patterson 1959, 1972; Caldeira and Patterson 1988), meaning not only will my results motivate a new line of congressional research, but the measure introduced in this study should help future scholars more easily measure complex social interactions, both on and off of Capitol Hill.

Acknowledgments

I am grateful to Robert X. Browning, Alan Cloutier, and the rest of the staff at the C-SPAN Archives for their help and support. I would also like to thank the anonymous reviewers and editor at *Political Analysis* as well as Arthur Spirling, Molly Roberts, Yiqing Xu, Jeff Mondak, and conference and seminar participants at UC-San Diego and PolMeth for all their helpful comments and suggestions. Finally, this paper would not have been possible without my fantastic research assistants: Jielu Yao and Logan Drake.

Data Availability Statement

The replication materials for this paper can be found at Harvard Dataverse at <https://doi.org/10.7910/DVN/YQPEVQ>.

Supplementary material

For supplementary material accompanying this paper, please visit <https://doi.org/10.1017/pan.2020.25>.

Bibliography

- Aggarwal, J. K., and Q. Cai. 1999. "Human Motion Analysis: A Review." *Computer Vision and Image Understanding* 73(3):428–440.
- Bogue, A. G., and M. P. Marlaire. 1975. "Of Mess and Men: The Boarding House and Congressional Voting, 1821–1842." *American Journal of Political Science* 45:207–230.
- Caldeira, G. A., and S. C. Patterson. 1987. "Political Friendship in the Legislature." *The Journal of Politics* 49(4):953–975.
- Caldeira, G. A., and S. C. Patterson. 1988. "Contours of Friendship and Respect in the Legislature." *American Politics Quarterly* 16(4):466–485.
- Casas, A., and N. W. Williams. 2019. "Images that Matter: Online Protests and the Mobilizing Role of Pictures." *Political Research Quarterly* 72(2):360–375.
- Cho, W. K. T., and J. H. Fowler. 2010. "Legislative Success in a Small World: Social Network Analysis and the Dynamics of Congressional Legislation." *The Journal of Politics* 72(1):124–135.
- Cohen, L., and C. J. Malloy. 2014. "Friends in High Places." *American Economic Journal: Economic Policy* 6(3):63–91.
- de Almeida, I. R., V. J. Cassol, N. I. Badler, S. R. Musse, and C. R. Jung. 2016. "Detection of Global and Local Motion Changes in Human Crowds." *IEEE Transactions on Circuits and Systems for Video Technology* 27(3):603–612.
- Dietrich, B. J. 2020. "Replication Data for: Using Motion Detection to Measure Social Polarization in the U.S. House of Representatives." <https://doi.org/10.7910/DVN/YQPEVQ>, Harvard Dataverse, V1.
- Dietrich, B. J., R. D. Enos, and M. Sen. 2019. "Emotional Arousal Predicts Voting on the Us Supreme Court." *Political Analysis* 27(2):237–243.
- Dietrich, B. J., M. Hayes, and D. Z. O'Brien. 2019. "Pitch Perfect: Vocal Pitch and the Emotional Intensity of Congressional Speech." *American Political Science Review* 113(4):941–962.
- Gibson, C. 2010. "Restoring Comity to Congress." Discussion Paper Series 60. Joan Shorenstein Center on the Press, Politics and Public Policy. <https://shorensteincenter.org/restoring-comity-to-congress/>.
- Green, A., and B. Hogan. 1982. "Gavel to Gavel: A guide to the Televised Proceedings of Congress." Technical report, Benton Foundation.
- Huddy, L., L. Mason, and L. Aarøe. 2015. "Expressive Partisanship: Campaign Involvement, Political Emotion, and Partisan Identity." *American Political Science Review* 109(1):1–17.
- Hughes, H. E. et al. 1976. "Toward a Modern Senate: Final Report of the Commission on the Operation of the Senate." Technical report, U.S. Senate.
- Kirkland, J. H. 2011. "The Relational Determinants of Legislative Outcomes: Strong and Weak Ties Between Legislators." *The Journal of Politics* 73(3):887–898.
- Koppensteiner, M., and K. Grammer. 2010. "Motion Patterns in Political Speech and Their Influence on Personality Ratings." *Journal of Research in Personality* 44(3):374–379.
- Koprinska, I., and S. Carrato. 2001. "Temporal Video Segmentation: A Survey." *Signal Processing: Image Communication* 16(5):477–500.
- Li, T., H. Chang, M. Wang, B. Ni, R. Hong, and S. Yan. 2014. "Crowded Scene Analysis: A Survey." *IEEE Transactions on Circuits and Systems for Video Technology* 25(3):367–386.
- Makin, D. A., D. W. Willits, W. Koslicki, R. Brooks, B. J. Dietrich, and R. L. Bailey. 2019. "Contextual Determinants of Observed Negative Emotional States in Police–Community Interactions." *Criminal Justice and Behavior* 46(2):301–318.
- Manchanda, S., and S. Sharma. 2016. "Analysis of Computer Vision Based Techniques for Motion Detection." In *2016 6th International Conference-Cloud System and Big Data Engineering (Confluence)*, 445–450. IEEE.
- Masket, S. E. 2008. "Where You Sit is Where You Stand: The Impact of Seating Proximity on Legislative Cue-Taking." *Quarterly Journal of Political Science* 3:301–311.
- Matthews, D. R. 1959. "The Folkways of the United States Senate: Conformity to Group Norms and Legislative Effectiveness." *American Political Science Review* 53(4):1064–1089.
- Murthy, G., and R. Jadon. 2009. "A Review of Vision Based Hand Gestures Recognition." *International Journal of Information Technology and Knowledge Management* 2(2):405–410.
- Mutz, D. C., and J. J. Mondak. 2006. "The Workplace as a Context for Cross-Cutting Political Discourse." *The Journal of Politics* 68(1):140–155.
- Patterson, S. C. 1959. "Patterns of Interpersonal Relations in a State Legislative Group: The Wisconsin Assembly." *Public Opinion Quarterly* 23(1):101–109.
- Patterson, S. C. 1972. "Party Opposition in the Legislature: The Ecology of Legislative Institutionalization." *Polity* 4(3):344–366.
- Paxton, A., and R. Dale. (2013). "Frame-Differencing Methods for Measuring Bodily Synchrony in Conversation." *Behavior Research Methods* 45(2), 329–343.
- Prajapati, D., and H. J. Galiyawala. 2015. "A Review on Moving Object Detection and Tracking." *International Journal of Computer Application* 5(3):168–175.

- Ramseyer, F., and W. Tschacher. 2014. "Nonverbal Synchrony of Head-and Body-Movement in Psychotherapy: Different Signals Have Different Associations With Outcome." *Frontiers in Psychology* 5:979.
- Seshadrinathan, K., and A. 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, 1-869–1-872. IEEE.
- Shafie, A. A., F. Hafiz, and M. Ali 2009. "Motion Detection Techniques Using Optical Flow." *World Academy of Science, Engineering and Technology* 56:559–561.
- Torres, M. 2018. "Give me the full picture: Using computer vision to understand visual frames and political communication." <http://qssi.psu.edu/new-faces-papers-2018/torres-computer-vision-and-politicalcommunication>.
- Valentino, N. A., T. Brader, E. W. Groenendyk, K. Gregorowicz, and V. L. Hutchings. 2011. "Election Night's Alright for Fighting: The Role of Emotions in Political Participation." *The Journal of Politics* 73(1):156–170.
- Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. 2004. "Image Quality Assessment: From Error Visibility to Structural Similarity." *IEEE Transactions on Image Processing* 13(4):600–612.
- Xi, N., D. Ma, M. Liou, Z. C. Steinert-Threlkeld, J. Anastasopoulos, and J. Joo. 2019. "Understanding the Political Ideology of Legislators From Social Media Images." *Proceedings of the International AAAI Conference on Web and Social Media* 14(1):726–737.
- Zauner, C. 2010. "Implementation and Benchmarking of Perceptual Image Hash Functions." Master's thesis, Upper Austria University of Applied Sciences.
- Zhong, Z., W. Ye, S. Wang, M. Yang, and Y. Xu. 2007. "Crowd Energy and Feature Analysis." In *2007 IEEE International Conference on Integration Technology*, 144–150. IEEE.