


Social Psychological Bulletin

Counter-Intuitive Findings on Affect and Ideology Likely Reflect Collider Bias: Commentary on Turner-Zwinkels et al., 2025

David J. Young¹ 

[1] *Department of Psychology, University of Cambridge, Cambridge, United Kingdom.*

Social Psychological Bulletin, 2026, Vol. 21, Article e15869, <https://doi.org/10.32872/spb.15869>

Received: 2024-10-18 • Accepted: 2025-06-13 • Published (VoR): 2026-03-17



Handling Editor: Roland Imhoff, Johannes Gutenberg University, Mainz, Germany

Corresponding Author: David J. Young, University of Cambridge, Department of Psychology, Downing Place, Cambridge, CB2 3EB, United Kingdom. E-mail: dy286@cam.ac.uk

Supplementary Materials: Code [see [Index of Supplementary Materials](#)]



Abstract

A paper recently published by Turner-Zwinkels et al. (2025) contains a peculiar set of findings: using cross-sectional analyses of international survey data, one study found that the more partisans agree with their co-partisans on matters of policy and ideology, the *less* they like the party they all support, whereas in another study these variables were uncorrelated. These negative and null relationships are counter-intuitive and conflict with prior findings showing that greater perceived ideological similarity tends to increase liking. Turner-Zwinkels et al. suggest these results could be a product of optimal distinctiveness theory. However, I suggest that this result is a statistical artefact caused by collider bias. I explain how collider bias could create this result even if the true relationship between liking and ideological similarity is positive. I demonstrate the plausibility of this explanation using simulations and a re-analysis of Turner-Zwinkels et al.'s data.

Keywords

Collider bias, Conditioning on a collider, Affective polarization, Ideological similarity, Policy positions, Partisanship



This is an open access article distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), CC BY 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Highlights

- A surprising recent results conflicts with a wealth of literature on the causes of affective polarization.
- Specifically, while much prior research shows that people prefer others who share their policy positions, [Turner-Zwinkels et al. \(2025\)](#) find, in two cross-sectional analyses, that partisans like their party more when they are more dissimilar to their co-partisans on matters of policy and ideology.
- I show that this result is most likely a statistical artefact caused by conditioning on a collider. I therefore raise greater awareness of collider bias and show that no reassessment of the literature with which the original finding conflicts is needed.

A large cross-disciplinary literature on affective polarization, i.e., mutual disliking between political groups ([Iyengar et al., 2012, 2019](#)), has developed in recent years, propelled by the observation that affective polarization has increased in the US in recent decades, a trend not widely seen in other developed nations ([Boxell et al., 2022](#)). Disagreement persists over the origins of affective polarization. Some scholars emphasize the role of *social identity* (e.g., [Dias & Lelkes, 2022](#)), arguing that partisans cherish their identities as supporters of a particular party, and this identification “triggers both positive feelings for the in group and negative evaluations of the out group” ([Iyengar et al., 2019](#), p.130), in line with classic Social Identity Theory (e.g., [Tajfel & Turner, 1979](#)). From this perspective, US affective polarization has increased due to *sorting*, with liberals becoming more likely to identify as Democrats and conservatives as Republicans ([Mason, 2015, 2018a, 2018b](#)).

An alternative view emphasises the role of *substantive* concerns, in particular policy disagreements; from this perspective, partisans dislike each other because they sincerely believe the policies their opponents support transgress moral principles or have undesirable practical consequences ([Bougher, 2017; Orr et al., 2023; Orr & Huber, 2020](#)). Indeed, many studies find that people prefer citizens, politicians, and parties when they perceive them to be similar to themselves in terms of ideology and policy preferences ([Algara & Zur, 2023; Bougher, 2017; Dias & Lelkes, 2022; Druckman et al., 2022; Lelkes, 2021; Orr et al., 2023; Orr & Huber, 2020; Rogowski & Sutherland, 2016; Webster & Abramowitz, 2017](#)). Notably, supporters of the identity account do not dispute these effects, rather, they suggest they occur because sharing policy preferences implies shared partisanship ([Dias & Lelkes, 2022](#)).

However, recent findings from [Turner-Zwinkels et al. \(2025\)](#) conflict with these results¹. They compared, in two large-*N* cross-national cross-sectional studies, how ideo-

1) Throughout this manuscript I refer to the results presented in Turner-Zwinkels et al.'s correction to their original paper (except where noted). In the course of writing this article, I noticed some coding errors in Turner-Zwinkels et

logically *similar* people are to other citizens who support the same party as them, and how much they *like* that party². In one study of eight European nations ($N = 4,246$), they found that similarity and liking are *negatively* correlated – the more similar people are to their co-partisans the *less* they like the party – and in a second study of 42 elections from 36 nations across the World ($N = 30,053$), they found no relationship.

These results are extremely surprising given prior literature, as experimental studies have shown that manipulating perceptions of the ideology of in-party politicians and citizens causes people to like them more when they are perceived to be more similar to themselves (Lelkes, 2021; Orr et al., 2023; Orr & Huber, 2020). In fact, these effects are quite large – Lelkes (2021) found that people with ideologically “extreme” views rated in-party politicians whose policy views were as “extreme” as their own 36 points higher than moderate politicians on a 0-100 feelings thermometer scale, and Orr et al. (2023) found that partisans rate co-partisans who share the same policy position as them 16 points higher than co-partisans who have positions they disagree with. Indeed, Turner-Zwinkels et al. originally hypothesized that they too would find positive correlations.

In their original article, Turner-Zwinkels et al. tentatively suggested that optimal distinctiveness theory (Brewer, 1991; Leonardelli et al., 2010), which proposes that people seek to balance cohering with a group against being distinct from other in-group members, might explain their negative correlation:

“[...] optimal distinctiveness theory could argue that people need more ingroup dissimilarity. As such, people should value the formation of subgroups within the ingroup to maintain greater individual distinctiveness.” (Turner-Zwinkels et al., 2025, p. 13)

However, while optimal distinctiveness theory might suggest that people like the *state of affairs* of being dissimilar to the rest of their party, it does not seem to follow that it would affect their liking *of the party itself*. Moreover, if this explanation were true, it is unclear why it would not have negatively biased affective ratings in the experimental literature where participants prefer more-similar in-party targets (Dias & Lelkes, 2022; Lelkes, 2021; Orr et al., 2023; Orr & Huber, 2020). Overall, this explanation seems unsatisfactory.

In this article, I argue that Turner-Zwinkels et al.’s findings can be best reconciled with prior experimental findings not by positing that there is some genuine difference in the causal mechanism that relates liking to ideological similarity when judging co-par-

al.’s original *R* scripts which meant that the party participants identified with was not always correctly identified; I notified the authors, prompting the correction.

2) These analyses comprise only a small amount of the many that Turner-Zwinkels et al. report, and almost the entirety of their article focuses on the theoretical implications of their numerous other findings. My focus on these results is because they are so striking and unexpected, rather than because they constitute a major focus of Turner-Zwinkels et al.’s article.

tisans compared to out-partisans, nor that prior methodologies were flawed or prior theorising mistaken. Rather, I suggest Turner-Zwinkels et al.'s findings are likely caused by collider bias. I support my argument with a re-analysis of the original data and through simulations (all code is available via <https://osf.io/skqub>).

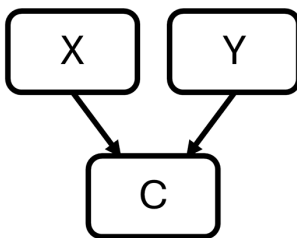
Collider Bias

Numerous guides to collider bias exist elsewhere—see [Pearl and Mackenzie \(2018\)](#), [Cinelli et al. \(2024\)](#), and [Lee et al. \(2019\)](#) for brief overviews, [Schneider \(2020\)](#) for a review of examples from economic history, and numerous cases from medical science ([Akimova et al., 2021](#); [Cole et al., 2010](#); [Hernán & Monge, 2023](#); [Weiskopf et al., 2023](#)). Additionally, a blogpost by Julia [Rohrer \(2017\)](#) discusses several highly accessible examples. Here, I will try to offer a brief, but intuitive explanation of what collider bias is and why it might matter in this case.

A collider is a variable whose value is influenced by two causally antecedent variables—see [Figure 1](#) for a diagram. The basic problem caused by collider bias is that if we control for a collider while attempting to measure the relationship between these two causally antecedent variables (or any variables they causally influence or are influenced by), we get a distorted estimate. Colliders can be contrasted with confounders, which are variables that also causally influence our dependent variable when we want to estimate the effect of a different independent variable on the dependent variable (i.e., in [Figure 1](#), Y is a confounder for our estimate of the effect of X on C), and mediators, which are intermediate variables on the causal path between our independent and dependent variables. Depending on our goals and the causal graph, controlling for confounders and mediators may or may not be appropriate (see [Cinelli et al., 2024](#)), but, except perhaps in very unusual circumstances, controlling for a collider is something we want to avoid.

Figure 1

Causal Graph Where C is a Collider for X and Y .



For example, the collider's value may be the sum of the other two variables. Let us keep calling the collider C and the other variables X and Y . Crucially, when C is a collider of X and Y , the value of X is not independent of the value of Y when the value of C is fixed, even if X and Y are otherwise independent. Therefore, any aspect of our data collection or analytical procedure which restricts the value of C will bias estimates of their relationship, as well as estimates of the relationships between variables that correlate with X and Y – generally speaking, collider bias can occur for any pair of variables which have a path between them that contains a collider (Cinelli et al., 2024). Such a restriction occurs if we control for C in a regression (Cinelli et al., 2024), or if inclusion within our dataset is conditioned on the value for C (e.g., Schneider, 2020). This latter case is known as *conditioning on a collider* (also known as *collider stratification bias*, *ascertainment bias*, and *Berkson's paradox*), and is the case of relevance for this article, because Turner-Zwinkels et al.'s analysis of whether the relationship between liking a party and being ideologically close to their supporters is moderated by supporting the party in question comprises two separate analyses, one where inclusion in the dataset is conditioned on the person supporting the relevant party, and one conditioned on them not.

To illustrate why conditioning on a collider is a problem, suppose $C = X + Y$, and we measure X , Y , and C across many cases, where X and Y are independent and have no correlation, with values in the range 0-100. If we restrict our dataset to only contain cases where $C = 100$, this will induce an artificial negative correlation between X and Y , because as X increases, Y must decrease in order for X and Y to sum to C : if X is 40, Y must be 60, if X is 80, Y must be 20, and so on. Thus, in this case, an artificial negative correlation occurs (and if $C = X - Y$, and we conditioned on, e.g., $C = 50$, an artificial positive correlation would be induced).

A similar bias emerges in less restrictive cases. Suppose we only specify $C > 80$. Again, this induces a negative bias in estimates of the correlation: if $X = 20$, Y must have values of 60-100, but if $X = 50$, Y must have values of 30-100, which will therefore tend to be lower on average, and if $X = 80$, Y must be in the even lower-bounded range 0-100, with an even lower average value. The lower bound on Y decreases as X increases, so as X goes up, the average value of Y will go down.

Note that the same also occurs when C is a *weighted* sum of X and Y , e.g., $C = w_X X + w_Y Y$. Suppose we condition on $C > k$. Then, for a given value of X , the lower bound on Y is $\frac{k - w_X X}{w_Y}$ (as $w_X X + w_Y Y > k$, so $w_Y Y > k - w_X X$, hence $Y > \frac{k - w_X X}{w_Y}$), which is a negative function of X . Therefore, as values of X increase, the average accompanying value of Y will tend to decrease, biasing the apparent correlation negatively.

Collider Bias in the Case of Turner-Zwinkels et al. (2025)

Conditioning on a collider may explain Turner-Zwinkels et al.'s negative and null correlations. Recall that these results come from Turner-Zwinkels et al.'s analysis of the relationship between the average similarity of a person's political belief to the supporters of a party ('Similarity') and how much they like that party on a 0–10 scale ('Liking'). To be a little more specific, Turner-Zwinkels et al. calculated 'Similarity' by finding the absolute mean distance between every pair of participants in each country/election-level sample across 10 (Study 1) or 9 (Study 2) items which measured policy preferences or ideological positioning (after first normalizing all variables to range from 0-1), then calculated how far each person was from each party's supporters on average, and rescaled so that high scores indicated greater proximity and low scores further distance.

Crucially, the surprising negative and null correlations are found using cases where *the participant supports the party that is the target of the Similarity and Liking ratings*. Yet, supporting a party is almost certainly a collider on the path between Similarity and Liking, as the more a person likes and is ideologically close to the supporters of a party, the more likely it is they will support that party. This occurs even if the decision to support a party is not *directly* influenced by ideological similarity to the *supporters* of that party, but rather by ideological similarity to the party's *policies* and/or elites, as this kind of similarity will tend to positively correlate with similarity to the party's supporters. Furthermore, note that the door remains open to collider bias even if we think there are other causal pathways at work, e.g., from party identification to liking and similarity, or from similarity to liking directly (see Cinelli et al., 2024).

I suggest a simple model can capture this scenario, whereby people are more likely to support a party if their summed Liking and Similarity score exceeds a relatively high threshold. This would be true if people prize both Liking and Similarity (either to the party's supporters, or to the party's elites and policy positions) when deciding who to support, but being high on one can compensate for being a little lower on the other, i.e., I can support a party even if there is some ideological distance between us if I like them for some other reason, or overlook some reasons to dislike the party if they are ideologically close to me. This sum must be high enough to warrant them choosing that party to support over others (as participants could only indicate support for one party), and high enough for them to warrant saying they support any party at all.

This creates a similar scenario to the example above where $C = X + Y$, with C being a continuous latent measure of the person's attraction towards the party, and X and Y being Similarity and Liking, but people only entering into the dataset if their C value is above the high threshold required for them to actually say they support them. Consequently, those high in Liking can 'get away' with having lower Similarity scores and yet still support the party (and vice versa), leading higher Liking scores to be found

alongside lower Similarity scores, on average. Even if Similarity and Liking normally have a positive correlation, this induces a negative bias, pushing it downwards.

Simulations and Re-Analysis

To test whether this proposal does provide a plausible counter-explanation for Turner-Zwinkels et al.'s surprising results, I conducted proof-of-concept simulations, supported by some re-analysis of both of Turner-Zwinkels et al.'s studies, which I performed using Turner-Zwinkels et al.'s description of their methodology, following it entirely, and successfully replicating their correlation estimates (from the Correction to the original article), as well as performing some new analyses. All scripts are available in the OSF project for this article.

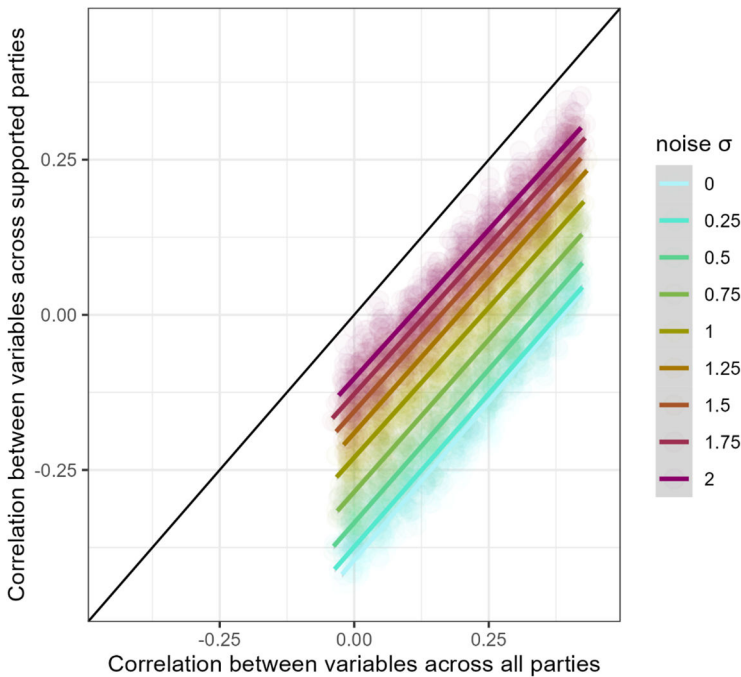
My first piece of re-analysis supports the general idea that this simple model is plausible, as participants' (equally-weighted) summed Liking and Similarity scores do exceed relatively high thresholds for their in-party. In Study 1, 88.2% of participants support the party for which they have the highest summed liking and similarity, with 84.1% doing so in Study 2. Additionally, the summed liking and similarity scores for supported parties fall, relative to people's summed liking and similarity scores for all rated parties, at the 96th percentile on average in Study 1 and the 95th percentile in Study 2, indicating that they exceed a high threshold.

My simulations show that with this model, a positive correlation between Similarity and Liking overall can turn null or negative when only in-party scores are analysed. I simulated 4050 electoral samples of 1000 agents each. Each agent has Similarity and Liking scores for 5 parties, and chooses to support the party with the highest sum for their Similarity score, Liking score, and the score from a random noise variable to capture, to some degree, the influence of unmodelled complexities in how people choose parties to support. Within each electoral sample, the 5000 Similarity scores are drawn at random from a normal distribution with a mean of 0.5 and a standard deviation of 1, which is also the case for the Liking scores; across simulations, I vary the correlation between Similarity and Liking from 0 to 0.4 in increments of 0.05. The noise variable is also normally distributed, with a mean of 0, and a standard deviation varied across simulations (as the 'noise σ ' parameter), from 0 to 2 in increments of 0.25—a higher standard deviation means the agent's choice of which party to support will be less influenced by Similarity and Liking. This creates $9 \times 9 = 81$ unique combinations of parameters, for each of which I generate 50 random samples, giving the 4050 electoral samples. For each sample, I calculate an "overall" correlation between Similarity and Liking across the agents' scores for all 5 parties, which should be equal to the programmed correlation plus random error. I also calculate an "in-party" correlation between Similarity and Liking using only the scores the agents have for the party they support.

Figure 2 shows the relationship between these correlations. Each point represents the pair of correlations obtained for one electoral sample. The x-axis shows the “overall” correlation between Similarity and Liking across scores for all parties for that simulation, and the y-axis shows the “in-party” correlation across the scores the agents give to the party they support. The points fall *below* the diagonal, demonstrating a downward bias of in-party correlations relative to overall correlations. This bias is stronger with less noise, but equal across different levels of overall correlation.

Figure 2

The Correlation Between Liking and Similarity Across All Parties (X-Axis) and for the Parties the Agents Support (Y-Axis) Across Simulations



Note. Linear regression lines are shown for each level of noise.

In concurrence with the results of these simulations, my analysis of Turner-Zwinkels et al.’s data finds that estimates of the correlation between Liking and Similarity are substantially biased downwards when only “in-party” data is used. For Study 1, an “overall” correlation of 0.376 [0.363, 0.389] ($p < .001$) drops to -0.096 [-0.126, -0.066] ($p < .001$) for the “in-party” correlation, and for Study 2, 0.216 [0.211, 0.221] ($p < .001$) drops to 0.005 [-0.007, 0.016] ($p = .424$). Further simulations in the Supplementary Materials

show that when liking and similarity scores are additionally influenced *by* which party the agents support, this worsens the bias.

Conclusion

The surprising negative and null relationships observed by Turner-Zwinkels et al. between liking the party you support and being ideologically similar to your co-partisans may be a statistical artefact, caused by conditioning on a collider. There is therefore probably no need to re-think existing findings, theories, or methodologies within the literature showing that people prefer those who concur with their ideology and policy positions (Algara & Zur, 2023; Bougher, 2017; Dias & Lelkes, 2022; Druckman et al., 2022; Lelkes, 2021; Orr et al., 2023; Orr & Huber, 2020; Rogowski & Sutherland, 2016; Webster & Abramowitz, 2017).

While I cannot definitively prove that collider bias explains Turner-Zwinkels et al.'s surprising results, the confluence of a) simulations which demonstrate the plausibility of this explanation in principle, and b) supporting statistical evidence from re-analyses of Turner-Zwinkels et al.'s data, makes a compelling case. Moreover, positing collider bias as the explanation provides a parsimonious account of how these results could arise when previous experimental studies suggest that ideological proximity to co-partisans boosts liking (Dias & Lelkes, 2022; Orr et al., 2023; Orr & Huber, 2020), as it avoids the need to theorise new causal mechanisms or identify methodological flaws in these studies.

This article highlights the particular danger collider bias poses for the interpretation of cross-sectional moderation analysis. Because collider bias can lead the estimated relationship between two variables to change *drastically* when a third variable is controlled for, even when that variable does not causally affect their relationship, it can give a misleading impression that the third variable acts as a moderator.³ This is plausibly what happened in Turner-Zwinkels et al., where the third variable was party identification, which appeared to moderate the relationship between liking a party and being ideologically close to their supporters, as the observed relationship was much stronger for parties the participants did not support than for those they did. Since the problem of collider bias is not widely appreciated in the context of moderation—indeed, Hayes' textbook companion to the widely-used PROCESS software does not mention colliders at all (Hayes, 2022)—researchers may waste resources exploring non-existent causal mechanisms without better awareness of collider bias. Researchers should therefore determine whether their moderators could be colliders (i.e., are causally influenced by both the IV and DV), consulting existing guides for doing so (Elwert & Winship, 2014; MacKinnon

3) It should also be noted that in other cases, collider bias could eliminate or reverse the appearance of 'true' causal moderation effects, giving a misleading impression about moderation in other ways.

& Lamp, 2021; Rohrer, 2018; Rohrer et al., 2022), and communicate explicitly about the limitations caused by possible collider bias where this cannot be ruled out, or better yet, use experimental methods where moderators can be manipulated, if possible. Reviewers too should be wary of the potential for cross-sectional moderation analyses to yield misleading results due to collider bias.

Overall, social and political psychologists should beware collider bias; it can cause artificial correlations to appear from nowhere, and for real correlations to be exaggerated, eliminated, or reversed. Therefore, when studying relationships between pairs of variables within sub-samples, such as partisans, if membership of the sub-sample could be conditional on both variables, it should be expected that cross-sectional estimates of the relationships will be distorted.

Funding: David Young's current position is funded by a Templeton World Charity Foundation grant, TWCF Number 31453.

Acknowledgments: Thanks to Felicity Turner-Zwinkels and Mark Brandt for a great degree of support and encouragement with this work; Julia Rohrer for very helpful comments; Lee de-Wit and Dave Lagnado for encouragement and useful discussions about this article; and to several anonymous reviewers for their constructive feedback.

Competing Interests: The author has declared that no competing interests exist.

Ethics Statement: Ethical approval was not required for this study, as it involved the re-analysis of existing data and did not include the collection of new data from human participants.

Data Availability: The data analysed belongs to third parties, which I therefore do not have permission to share. But I share my code and download links to the data are available in my OSF project (see Young, 2024).

Supplementary Materials

For this article, code and download links to the data are available (see Young, 2024).

Index of Supplementary Materials

Young, D. J. (2024). *Counter-intuitive findings on affect and ideology likely reflect collider bias: Commentary on Turner-Zwinkels et al., 2023*. [Code, link to original data]. OSF.
<https://osf.io/skqub>

References

- Akimova, E. T., Breen, R., Brazel, D. M., & Mills, M. C. (2021). Gene-environment dependencies lead to collider bias in models with polygenic scores. *Scientific Reports*, *11*(1), Article 9457. <https://doi.org/10.1038/s41598-021-89020-x>
- Algara, C., & Zur, R. (2023). The Downsian roots of affective polarization. *Electoral Studies*, *82*, Article 102581. <https://doi.org/10.1016/j.electstud.2023.102581>
- Bougher, L. D. (2017). The correlates of discord: Identity, issue alignment, and political hostility in polarized America. *Political Behavior*, *39*(3), 731–762. <https://doi.org/10.1007/s11109-016-9377-1>
- Boxell, L., Gentzkow, M., & Shapiro, J. M. (2022). Cross-country trends in affective polarization. *The Review of Economics and Statistics*, *106*(2), 557–565. https://doi.org/10.1162/rest_a_01160
- Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality & Social Psychology Bulletin*, *17*(5), 475–482. <https://doi.org/10.1177/0146167291175001>
- Cinelli, C., Forney, A., & Pearl, J. (2024). A crash course in good and bad controls. *Sociological Methods & Research*, *53*(3), 1071–1104. <https://doi.org/10.1177/00491241221099552>
- Cole, S. R., Platt, R. W., Schisterman, E. F., Chu, H., Westreich, D., Richardson, D., & Poole, C. (2010). Illustrating bias due to conditioning on a collider. *International Journal of Epidemiology*, *39*(2), 417–420. <https://doi.org/10.1093/ije/dyp334>
- Dias, N., & Lelkes, Y. (2022). The nature of affective polarization: Disentangling policy disagreement from Partisan identity. *American Journal of Political Science*, *66*(3), 775–790. <https://doi.org/10.1111/ajps.12628>
- Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M., & Ryan, J. B. (2022). (Mis-)estimating affective polarization. *The Journal of Politics*, *84*(2). <https://doi.org/10.1086/715603>
- Elwert, F., & Winship, C. (2014). Endogenous selection bias: the problem of conditioning on a collider variable. *Annual Review of Sociology*, *40*, 31–53. <https://doi.org/10.1146/annurev-soc-071913-043455>
- Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (3rd ed.). The Guilford Press.
- Hernán, M. A., & Monge, S. (2023). Selection bias due to conditioning on a collider. *BMJ (Clinical Research Ed.)*, *381*, Article 1135. <https://doi.org/10.1136/bmj.p1135>
- Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, *22*(1), 129–146. <https://doi.org/10.1146/annurev-polisci-051117-073034>
- Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology. *Public Opinion Quarterly*, *76*(3), 405–431. <https://doi.org/10.1093/poq/nfs038>
- Lee, H., Aronson, J. K., & Nunan, D. (2019). *Collider bias*. Catalogue of bias. <https://catalogofbias.org/biases/collider-bias/>
- Lelkes, Y. (2021). Policy over party: Comparing the effects of candidate ideology and party on affective polarization. *Political Science Research and Methods*, *9*(1), 189–196. <https://doi.org/10.1017/psrm.2019.18>

- Leonardelli, G. J., Pickett, C. L., & Brewer, M. B. (2010). Optimal distinctiveness theory. In M. P. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology* (Vol. 43, pp. 63–113). Academic Press. [https://doi.org/10.1016/S0065-2601\(10\)43002-6](https://doi.org/10.1016/S0065-2601(10)43002-6)
- MacKinnon, D. P., & Lamp, S. J. (2021). A unification of mediator, confounder, and collider effects. *Prevention Science*, 22, 1185–1193. <https://doi.org/10.1007/s11121-021-01268-x>
- Mason, L. (2015). “I disrespectfully agree”: The differential effects of partisan sorting on social and issue polarization: Partisan sorting and polarization. *American Journal of Political Science*, 59(1), 128–145. <https://doi.org/10.1111/ajps.12089>
- Mason, L. (2018a). Ideologues without issues: The polarizing consequences of ideological identities. *Public Opinion Quarterly*, 82(S1), 866–887. <https://doi.org/10.1093/poq/nfy005>
- Mason, L. (2018b). *Uncivil agreement: How politics became our identity*. University of Chicago Press.
- Orr, L. V., Fowler, A., & Huber, G. A. (2023). Is affective polarization driven by identity, loyalty, or substance? *American Journal of Political Science*, 67(4), 948–962. <https://doi.org/10.1111/ajps.12796>
- Orr, L. V., & Huber, G. A. (2020). The policy basis of measured partisan animosity in the United States. *American Journal of Political Science*, 64(3), 569–586. <https://doi.org/10.1111/ajps.12498>
- Pearl, J., & Mackenzie, D. (2018). *The book of why*. Penguin Random House.
- Rogowski, J. C., & Sutherland, J. L. (2016). How ideology fuels affective polarization. *Political Behavior*, 38(2), 485–508. <https://doi.org/10.1007/s11109-015-9323-7>
- Rohrer, J. M. (2017). *That one weird third variable problem nobody ever mentions: Conditioning on a collider*. The 100% CI. <https://www.the100.ci/2017/03/14/that-one-weird-third-variable-problem-nobody-ever-mentions-conditioning-on-a-collider/>
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1(1), 27–42. <https://doi.org/10.1177/2515245917745629>
- Rohrer, J. M., Hünemann, P., Arslan, R. C., & Elson, M. (2022). That’s a lot to process! Pitfalls of popular path models. *Advances in Methods and Practices in Psychological Science*, 5(2). <https://doi.org/10.1177/25152459221095827>
- Schneider, E. B. (2020). Collider bias in economic history research. *Explorations in Economic History*, 78, Article 101356. <https://doi.org/10.1016/j.eeh.2020.101356>
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin, & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–37). Brooks/Cole.
- Turner-Zwinkels, F. M., van Noord, J., Kesberg, R., García-Sánchez, E., Brandt, M. J., Kuppens, T., Easterbrook, M. J., Smets, L., Gorska, P., Marchlewska, M., & Turner-Zwinkels, T. (2025). Affective polarization and political belief systems: The role of political identity and the content and structure of political beliefs. *Personality & Social Psychology Bulletin*, 51(2), 222–238. <https://doi.org/10.1177/01461672231183935>

Webster, S. W., & Abramowitz, A. I. (2017). the ideological foundations of affective polarization in the U.S. electorate. *American Politics Research*, 45(4), 621–647.

<https://doi.org/10.1177/1532673X17703132>

Weiskopf, N. G., Dorr, D. A., Jackson, C., Lehmann, H. P., & Thompson, C. A. (2023). Healthcare utilization is a collider: An introduction to collider bias in EHR data reuse. *Journal of the American Medical Informatics Association*, 30(5), 971–977. <https://doi.org/10.1093/jamia/ocad013>

PSPS Polish
Social
Psychological
Society

Social Psychological Bulletin (SPB) is an official journal of the Polish Social Psychological Society (PSPS).

zpid ^{LP}

Leibniz-Institut für
Psychologie

PsychOpen GOLD is a publishing service provided by the Leibniz Institute for Psychology (ZPID), Germany.