Measuring Public Opinion: "The Wasserstein Bipolarization Index", with Application to Cross-National Attitudes Toward Mandatory Vaccination for COVID-19

Hane Lee* Michael E. Sobel

Department of Statistics, Columbia University

Draft: October 18, 2024

Abstract

The extent to which the American public is politically polarized is of great interest in the lay and academic communities. To study opinion polarization, political scientists and public opinion researchers examine the distribution of respondents on survey items, using visual comparison of histograms, and/or measures such as variances and bimodality coefficients. We prove these measures fail to align with the conceptualization of polarization put forth in the literature. To remedy this situation we specify several properties a measure of polarization consistent with this conceptualization should possess: in particular, it should increase as a distribution spreads away from a center toward the poles and/or as clustering below or above this center increases. We then propose a p-Wasserstein bipolarization index that satisfies these properties and which measures the distance between the distribution of an item and a most polarized distribution with all mass concentrated on the lower and upper endpoints of the scale, using it to examine bipolarization in attitudes toward governmental COVID-19 vaccine mandates across 11 countries: the US and UK are most polarized, China, France, and India the least polarized, with Spain, Colombia, Italy, Brazil, Australia and Canada occupying an intermediate position.

Keywords: Polarization, Wasserstein Distance, Public Opinion, COVID-19 Vaccination Mandates.

*Email: hane.lee@columbia.edu

1 Introduction

Contemporary media coverage of the American political scene depicts a public sharply divided by different views into opposing camps whose members distrust and dislike those outside their cluster.

Political scientists have offered a more nuanced account. First, they distinguish between political elites and the public, widely agreeing that "elite polarization", i.e., ideological homogeneity among party representatives and ideological dissimilarity between representatives of different parties, has steadily increased since the 1970s (Poole and Rosenthal 2011). At the mass level, there is also a consensus that "partisan sorting", whereby individuals who self-identify as liberal (conservative) are likely to identify as Democrats (Republicans), has increased (Levendusky 2009; Fiorina and Abrams 2008). And there is also agreement that "affective polarization", whereby members or supporters of each party view those of the other party negatively, has increased. (Druckman and Levy 2022; Iyengar et al. 2019; Simas, Clifford, and Kirkland 2020).

Another strand of literature is concerned with "ideological polarization", the dissimilarity between members of the public on policy issues. Two notions have been espoused (Lelkes 2016). "Ideological consistency" refers to the extent to which persons espouse liberal or conservative opinions on multiple issues, whereas "ideological divergence" refers to the distribution of public opinion on given issues, with clustering of respondents at opposing extremes of a univariate scale indicative of polarization on that issue.

Abramowitz and Saunders (2008) equate polarization with ideological consistency (alignment). Using opinion items from the National Election Study (NES) they argue that consistency increased over the period 1982-2004, as evidenced by both the proportion of respondents giving consistently liberal or conservative responses to the items and the correlation of responses to different items. Similarly, the Pew Research Center (2014) reports that the percentage of persons who express ideologically consistent opinions increased from 1994-2014. However, using items from the 1972-2004 NES, Baldassarri and Gelman (2008) find little evidence of increased alignment. Similarly, using items from the General Social Survey (GSS) and the NES, DiMaggio, Evans, and Bryson (1996) find little evidence of increased "ideological constraint" in their analysis spanning 1972-1994. But a more recent analysis using NES items (Kozlowski and Murphy 2021) points to an increase in correlations among items over from 2004-2016. And in the Canadian context, Merkley (2022) finds that the average correlation between eight issue items from the Canadian Election Studies increased moderately from 1993-2019.

There is widespread agreement that ideological divergence is maximally manifested when individuals are concentrated in two clusters at opposite ends of a univariate ordinal or bounded metric scale (Bauer 2019), suggesting ideological divergence should be measured by comparing the observed distribution of opinion to this exemplar distribution. But researchers have taken a more elliptical approach: using visualization (Fiorina and Levendusky 2006; Fiorina and Abrams 2008; Koudenburg, Kiers, and Kashima 2021) and/or ad-hoc measures such as the sample variance and bimodality coefficients (DiMaggio, Evans, and Bryson 1996; Strijbis, Helmer, and De Wilde 2020; Melki and Pickering 2014; Lelkes 2016; Merkley 2022) that appear on the face of it to capture important features of the exemplar, they ask instead whether there is a temporal trend in these measures. For example, Fiorina and Abrams (2008) display figures of distributions intended to help readers visualize polarized vs. non-polarized distributions and they compare frequency distributions of five NES seven point ordinal opinion items in 1984 and 2004: for four of these, the percentage of respondents in the first, seventh and "central" categories appears similar at the two time points, indicating to them no change in polarization over this period. But they do not attempt to say whether or not any of the distributions exhibit polarization, stating that whether or not opinions are polarized at a particular time point "is generally a matter of judgement" (p.566). DiMaggio, Evans, and Bryson (1996) argue that increasing the spread and bimodality of a distribution increases ideological divergence. But they do not attempt to formalize and connect these concepts to changes in polarization, simply operationalizing spread by variance, bimodality by kurtosis; Mouw and Sobel (2001) point out that kurtosis does not measure bimodality. In their empirical work, DiMaggio et al. find no evidence (with the exception of an item measuring opinions about abortion) that attitudes have polarized. Hill and Tausanovitch (2015) use 67 NES policy items collected in different years between 1956 and 2012, reporting no increase in spread as measured by the variance of estimated ideal points for each respondent. Using NES items from 1972-2012 and Sarle's sample bimodality coefficent, Lelkes (2016) finds no evidence of increasing polarization, and using NES items from 1972-2016, Fiorina (2017) comes to the same conclusion.

Economists have also considered polarization. Here the aforementioned disconnect between measurement and conceptualization is bridged through an axiomatic approach in which intuitions about polarization are formalized, and measures conforming to the axioms constructed. Two strands of literature predominate (Esteban and Ray 2012). The first (Esteban and Ray 1994; Duclos, Esteban, and Ray 2004; Esteban, Gradín, and Ray 2007) views polarization as the grouping of individuals into several (not necessarily two) clusters where members of different groups are possibly antagonistic toward members of other clusters: polarization is then the sum of all these antagonisms. The second strand, which more closely informs our research, stems from work on the "hollowing out" of the "middle class" and views polarization, (synonymously bipolarization in this context) as the clustering of individuals on opposite sides of the median of a metric scale (e.g, income), with greater distances from the median indicating higher levels of polarization. Here the notions of spread and bipolarity are axiomatized, leading, in conjunction with functional form and normalization restrictions, to indices that essentially measure the dissimilarity between an observed distribution and a distribution concentrated at the median of the observed distribution. To the best of our knowledge, statistical properties of these indices have not been explored. See Anderson (2004), who uses stochastic dominance relations to characterize and test for various forms of polarization.

There is also a related literature on bipolarization for ordinal scales (Allison and Foster 2004; Apouey 2007; Apouey and Silber 2013; Kobus 2015; Sarkar and Santra 2020). In future work, we shall address this case; for further remarks, see the discussion.

In section two, axioms that we want an index of polarization to satisfy are set forth. To measure the distance between an observed distribution of opinion and a maximally polarized distribution, we propose using the p-Wasserstein distance (a measure of the distance between two probability distributions defined for $p \geq 1$) and show it is consistent with our axioms. We also show that commonly used measures such as the variance and Sarle's sample bimodality coeffĭcent are incompatible with one or more of the axioms. In section three, for $p \in \{1, 2\}$, the estimated p-Wasserstein index is used to evaluate bipolarization among 11 countries in attitudes toward mandatory COVID-19 vaccination. The US and UK are the most polarized countries, China, India and France the least, while Brazil, Australia, Colombia, Canada, Italy and Spain are intermediate. Section four concludes. The appendix describes the statistical theory on which our interval estimates of polarization are based. An R package (WBI: Wasserstein Bipolarization Index) was developed to carry out the computations.

2 Bipolarization: A New Measure

We consider responses X, with associated probability measure ν_X, to opinion item $\mathcal{I}$ measured on a metric scale with compact support on $S \subseteq [\ell, L] \subset \mathbb{R}^1$. We propose five axioms, two of which formalize the notion that bipolarization increases as the spread and clustering of responses on either side of a central point c increase. Our treatment is closely related to that of Foster and Wolfson (2010) and Wang and Tsui (2000), whose indices measure the dissimilarity between ν_X and a measure concentrated at the median $m(\nu_X)$ of ν_X. In contrast, our index measures the distance of ν_X, with center c (not necessarily equal to $m(\nu_X)$) from a maximally separated measure $\xi_{(\nu_X,c,\gamma)}$ concentrated at the lower and upper poles ℓ and L.

Definition. The Maximally Separated Measure. For $X \in [\ell, L]$, with probability measure ν_X, the maximally separated measure $\xi_{(\nu_X,c,\gamma)}$, with center $c \in (\ell, L)$, assigns all probability to the points ℓ and L as follows: $\xi_{(\nu_X,c,\gamma)}(\ell) = \nu_X[\ell, c) + \gamma\nu_X(c)$, $\xi_{(\nu_X,c,\gamma)}(L) = 1 - \xi_{(\nu_X,c,\gamma)}(\ell)$, where $\gamma \in [0, 1]$.

For the case $\nu_X(c) = 0$, for any γ, the maximally separated measure results from transferring all the mass of ν_X to the left of center to ℓ and all the mass to the right of center to L. When $\nu_X(c) > 0$, the mass to the left (right) of c is transferred to ℓ (L), and a proportion γ, specified by the researcher, of the mass at the center is transferred to ℓ, the remainder to L. If c is chosen to correspond to a quantile q of ν_X, γ is chosen so that the maximally separated measure places mass q on ℓ and mass $1 - q$ on L.

In public opinion research the variance is often used to measure bipolarization, in which case the center is the mean. Substantively motivated choices are also possible. For items where a point on the scale represents a neutral position, a researcher might want to choose this point as a center. Public opinion researchers widely agree that the maximum amount of bipolarization occurs when half the responses to an item are clustered at the lower endpoint, the other half clustered at the upper endpoint (Bauer 2019). This implies c is the median. The median also figures prominently in the political science literature, where the "average citizen" often refers to the median voter (Gilens and Page 2014; Downs 1957), and the median position historically indicates a "central tendency among voters" (Kim and Fording 2003). Similarly, in the economic literature on bipolarization, motivated by concerns over the "hollowing out of the middle class, the median income represents the center of this class.

It is important to understand the empirical implications of different choices of the center. If the center is chosen to correspond to a quantile q, i.e., the center of ν_X is $c(q)$, $\xi_{(\nu_X,c(q),\gamma(q))}$ depends only on ν_X and q. But if c is chosen in some other fashion, e.g., as the mean or some point representing a neutral position, $\xi_{(\nu_X,c,\gamma)}$ depends on all three parameters.

In section 2.1, we propose five axioms we want our index of bipolarization to satisfy. In section 2.2, we show that measures of polarization commonly used in the literature on ideological divergence are incompatible with one or more of these axioms. In section 2.3, we propose a p-Wasserstein measure of bipolarization and prove it is consistent with the five axioms.

2.1 Axioms

Axioms 1 and 2 impose invariance conditions on a bipolarization index P that facilitate comparison across populations and scales.

Axiom A1. Let X be a random variable with probability measure ν_X and Y a random variable with probability measure ν_Y. Then, if $\nu_X = \nu_Y$, $P(\nu_X, \xi_{(\nu_X,c,\gamma)}) = P(\nu_Y, \xi_{(\nu_X,c,\gamma)})$ for any choice of c and γ.

Second, the origin ℓ and maximum L of items measuring public opinion are arbitrary. For example, the NES feeling thermometer items range from "very cold or unfavorable feeling" ($\ell = 0$) to "very warm or favorable feeling" ($L = 100$); political ideology is often assessed on a 10 or 11 point scale running from "left" to "right" (Bauer et al. 2017). To facilitate interpretation and enable comparison of responses to one or more items measured on different scales, e.g., the same question asked in different surveys, or to ask whether attitudes toward gun control are more polarized than attitudes toward abortion, we require P to be translation invariant and homogeneous of degree 1.

Axiom A2. Let $X \in [\ell, L]$ denote the response to item $\mathcal{I}$ in population $\mathcal{P}$. Let $Y = \ell_Y + \beta(X - \ell) \in [\ell_Y, L_Y]$, $\beta = \frac{L_Y - \ell_Y}{L - \ell}$. Then $P(\nu_Y, \xi_{(\nu_Y, \ell_Y + \beta(c-\ell), \gamma)}) = \beta P(\nu_X, \xi_{(\nu_X,c,\gamma)})$.

Next, as in the economic literature on bipolarization (Foster and Wolfson 2010, Wang and Tsui 2000, Esteban and Ray 2012) of income, we formalize the notions of spread and clustering that are also commonly identified in the literature on ideological divergence as features of polarization. Motivated by earlier work on the "hollowing out of the middle class", Foster and Wolfson (2010) characterize increased spread as follows: for responses X and Y measured in the same way, with probability measures ν_X and ν_Y respectively, and common median $m(\nu_X) = m(\nu_Y) = m$, ν_Y is more polarized than ν_X if the distribution function F_Y is stochastically higher than F_X to the left of m and F_X is stochastically higher than F_Y to the right of m. The intuition for this definition is that movements from the left (right) of the median further left (right) increase polarization: below we characterize increased spread in terms of such movements.

Definition. Left spread. ν_Y is a left spread of ν_X about c if $\nu_Y(I) = \nu_X(I)$ for every sub-interval I of $(c, L]$, and there exists $x_1 \in [\ell, c]$ such that $\nu_Y(I) \geq \nu_X(I)$ for every sub-interval I of $[\ell, x_1)$ and $\nu_Y(I) \leq \nu_X(I)$ for every sub-interval I of $(x_1, c]$. A left spread about c is said to be restricted if $\nu_Y(c) \geq (1-\gamma)\nu_X(c)$. **Right Spread**. ν_Y is a right spread of ν_X about c if $\nu_Y(I) = \nu_X(I)$ for every sub-interval I of $[\ell, c)$, and there exists $x_1 \in [c, L]$ such that $\nu_Y(I) \leq \nu_X(I)$ for every sub-interval I of $[c, x_1)$ and $\nu_Y(I) \geq \nu_X(I)$ for every sub-interval I of $(x_1, L]$. A right spread about c is said to be restricted if $\nu_Y(c) \geq \gamma\nu_X(c)$.

The definitions above include the case of a null spread $\nu_X = \nu_Y$, in which case x_1 can be any value in $[\ell, c]$ for a left spread about c, any value in $[c, L]$ for a right spread about c. For a non-null left spread about c, with $x_1 \in (\ell, c)$, either mass is transferred from $(x_1, c]$ to $[\ell, x_1]$ or from $[x_1, c]$ to $[\ell, x_1)$; if $x_1 = \ell$, x_1 receives a transfer of mass from $(\ell, c]$, and if $x_1 = c$, mass is sent from x_1 toward ℓ. Similarly, for a non-null right spread about c, with $x_1 \in (c, L)$, either mass is transferred from $[c, x_1]$ to $(x_1, L]$ or from $[c, x_1)$ to $[x_1, L]$; if $x_1 = c$, x_1 sends mass toward L, and if $x_1 = L$, x_1 receives mass from $[c, L)$.

Axiom A3. If ν_Y is a restricted left (restricted right) spread of ν_X about c, $P(\nu_Y, \xi_{(\nu_X,c,\gamma)}) \leq P(\nu_X, \xi_{(\nu_X,c,\gamma)})$.

The restriction $\nu_Y(c) \geq (1-\gamma)\nu_X(c)$ on a left spread about c ensures that at most $\gamma\nu_X(c)$ probability mass is moved from c toward ℓ; otherwise, it would be necessary to transfer some mass in $[\ell, c)$ to c in order to attain the distribution corresponding to the maximally separated measure. Similarly, the restriction $\nu_Y(c) \geq \gamma\nu_X(c)$ on a right spread about c ensures that at most $(1-\gamma)\nu_X(c)$ probability mass is moved from c toward L.

A second feature of bipolarization is clustering: moving masses on the left (right) of center closer together should increase polarization. But this may also increase spread. Consider the case of a discrete distribution ν_X, where $\nu_X(x_1) = p_1 > 0$, $x_1 < x_2 \leq c$, $\nu_X(x_2) = p_2 > 0$. A new measure ν_Y is created by moving 1) mass $p_2^* < p_2$ from x_2 to $x_2 - \delta$ and 2) mass $p_1^* < p_1$ from x_1 to $x_1 + \eta$, where $x_1 < x_1 + \eta \leq x_2 - \delta < x_2$. Clearly, move 1 increases spread and move 2 decreases spread, and p_1^*, p_2^*, δ, and η may be chosen so that the net spread either increases or decreases. To capture changes in clustering independently of spread, these quantities should be chosen so that the net change in spread is 0, i.e., the mean value $E(Y)$ associated with ν_Y equals the mean value $E(X)$ associated with ν_X. To formalize this requirement, we first define a "mean preserving merge", then adapt this to the case of clustering on either side of c. Our definition is adapted from the definition of a mean preserving spread in Machina and Pratt (1997; 105-106), who generalized the definition of a mean preserving spread in Rothschild and Stiglitz (1970).

Definition. Mean preserving merge. Let ν_X and ν_Y be probability distributions on $[\ell, L]$ with common mean $E(X) = E(Y)$. We say ν_Y is obtained from ν_X by a mean preserving merge (equivalently, ν_X is obtained from ν_Y by a mean preserving spread) if there exist values $x_1 \leq x_2$ such that : a) $\nu_Y(I) \leq \nu_X(I)$ for every sub-interval I of $[\ell, x_1)$, b) $\nu_Y(I) \geq \nu_X(I)$ for every sub-interval I of (x_1, x_2), c) $\nu_Y(I) \leq \nu_X(I)$ for every sub-interval I of $(x_2, L]$.

The case $x_1 = x_2$ occurs for a null merge $\nu_X = \nu_Y$ or if mass is sent from $[\ell, x_1)$ and $(x_1, L]$ to x_1. The point x_1 can transfer probability to or from $\nu_X(x_1, x_2)$ and the point x_2 can transfer probability to or from $\nu_X(x_2, L]$. The Pigou-Dalton transfer, used by Wang and Tsui (2002) to characterize clustering, whereby an individual with income $x_1 > x_0$

transfers amount $\lambda < x_1 - x_0$ to the individual with income x_0, is a special case of the mean preserving merge. We now adapt the definition above to characterize clustering, in which masses "below" or "above" c are moved closer together.

Definition. Left Merge. Let $\nu_Y(I) = \nu_X(I)$ for every sub-interval I of $(c, L]$. If ν_Y is a mean preserving merge of ν_X with $x_2 \leq c$, we say ν_Y is a left mean preserving merge about c. We say a left mean preserving merge about c is restricted if $\nu_Y(c) \geq (1-\gamma)\nu_X(c)$. **Right Merge**. Let $\nu_Y(I) = \nu_X(I)$ for every sub-interval I of $[\ell, c)$. If ν_Y is a mean preserving merge of ν_X with $x_1 \geq c$, we say ν_Y is a right mean preserving merge about c. We say a right mean preserving merge about c is restricted if $\nu_Y(c) \geq \gamma\nu_X(c)$.

Axiom A4. If ν_Y is obtained from ν_X by a restricted left mean preserving merge about c or a restricted right mean preserving merge about c, $P(\nu_Y, \xi_{(\nu_X,c,\gamma)}) \leq P(\nu_X, \xi_{(\nu_X,c,\gamma)})$.

Our goal is to develop an index that measures the dissimilarity between ν_X and its maximally separated counterpart $\xi_{(\nu_X,c,\gamma)}$ in which the mass to the "left" of center is transferred to l and mass to the "right" to L. We require our measure to be a distance:

Axiom A5. For any two measures ν_X and ν_Y, $P(\nu_X, \nu_Y)$ is a distance.

2.2 Polarization Indices in Public Opinion Research: Variance and Bimodality

Public opinion researchers typically use either the variance or Sarle's sample bimodality coefficient to measure polarization, interpreting increases in these measures as evidence of increasing polarization. But neither of these measures are consistent with the axioms we have set forth. First, recall that a mean preserving merge is the reverse of a mean preserving spread, and it is well known that a mean preserving spread increases the variance. Therefore a mean preserving merge, i.e., increased clustering, reduces variance: thus, using the variance as a measure of polarization can lead to the inference, in conflict with Axiom A4, that bipolarization increases as clustering decreases. Second, consider Sarle's finite sample bimodality coefficient (Lelkes 2016):

$$b = \frac{g^2 + 1}{k + \frac{3(n-1)^2}{(n-2)(n-3)}}, \tag{1}$$

where $g = \frac{m_3}{s^3}$ is the sample skewness, $k = \frac{m_4}{s^4} - 3$ is the sample excess kurtosis, m_3 and m_4 are the sample third and fourth central moments, and s is the sample standard deviation. Clearly, Axiom A1 is violated, as two samples with different sizes and the same values of k and g yield different values of b. While this is inconsequential for "large" n, consider next two samples of size n with observed probability measures $\nu_X = 0.5 \cdot \delta_{0.4} + 0.5 \cdot \delta_{0.6}$ and $\nu_Y = 0.5 \cdot \delta_{0.1} + 0.5 \cdot \delta_{0.9}$ on $[0, 1]$, where δ_x denotes a Dirac measure with mass 1 at x. Clearly, ν_Y exhibits increased spread compared to ν_X, but g and k are equal, and b will suggest, at least in large samples, that ν_X and ν_Y are equally polarized, violating Axiom A3.

2.3 p-Wasserstein measure of polarization

We propose to measure the distance between ν_X and $\xi_{(\nu_X, c, \gamma)}$ using the p-Wasserstein distance, and we show that axioms A1-A5 are consistent with this choice. Our definition is adapted from Villani (2009):

Definition. p-Wasserstein distance. Let $(\mathcal{X}, d)$ be a Polish metric space, and let $p \in$

$[1, \infty)$. For any two probability measures ν_X, ν_Y on $\mathcal{X}$, the Wasserstein distance of order p is defined by

$$W_p(\nu_X, \nu_Y) = \left(\inf_{\pi \in \Pi(\nu_X, \nu_Y)} \int_{\mathcal{X}} d(x, y)^p d\pi(x, y) \right)^{\frac{1}{p}}, \qquad (2)$$

where $\Pi(\nu_X, \nu_Y)$ is the set of all joint probability measures on $\mathcal{X} \times \mathcal{X}$ with marginals ν_X and ν_Y, respectively.

For a given metric d, (2) is the minimum cost of "movement" needed to transform the distribution ν_X into ν_Y or ν_Y into ν_X (Solomon 2018). Here we take $d(x, y) = |y - x|$.

Our choice of the Wasserstein distance is motivated by it capacity to capture "key geometric properties of the underlying ground space" that other statistical distances do not (Peyré and Cuturi 2019). Consider Figure 1 where ν_X is comprised of masses A and B, with c equal to the median $m(\nu_X)$, and denote the most polarized distribution, with masses .5 at ℓ and L, ξ_{pol}. Let ν_Y be obtained from ν_X by shifting B k units to the right (a right spread) to B'. Intuitively, ν_Y exhibits greater polarization than ν_X: further, to satisfy axiom A3 our measure $D(\cdot, \cdot)$ must satisfy $D(\nu_Y, \xi_{\text{pol}}) < D(\nu_X, \xi_{\text{pol}})$, as will be the case using the p-Wasserstein distance. However, other commonly used distances, such as the total variation distance between probability measures ν_X and ν_Y ($D_{\text{TV}}(\nu_X, \nu_Y) := \sup_{S \subset \mathcal{X}} |\nu_X(S) - \nu_Y(S)|$), and the Hellinger distance ($D_{\text{H}}(\nu_X, \nu_Y) := \frac{1}{2} \int_{\mathcal{X}} (\sqrt{\nu_X(dx)} - \sqrt{\nu_Y(dx)})^2$) will not discriminate between these cases because neither ν_X nor ν_Y share any points of common support with ξ_{pol}: nor do these distances satisfy Axioms A2-A4. And the Kullback-Leibler (KL) divergence ($D_{\text{KL}}(\nu_X, \nu_Y) := \int_{\mathcal{X}} \log \left(\frac{\nu_X(dx)}{\nu_Y(dx)} \right) \nu_X(dx)$), another popular measure of the dissimilarity between distributions, albeit not a distance, is not applicable, as ν_X (ν_Y) and ξ_{pol} do not share a common support.

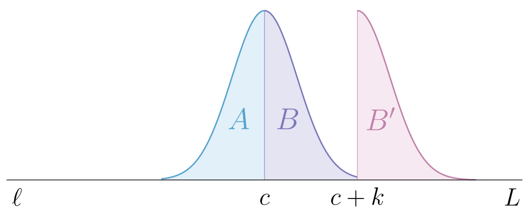

Figure 1: ν_X consists of masses A and B, and ν_Y consists of masses A and B', where the mass B' results from translating the mass B k units to the right. ν_Y is more polarized than ν_X.

Calculating the Wasserstein distance requires finding the optimal coupling between two measures that minimizes the cost of transport. Suppose for the moment $\nu_X(c) = 0$. It seems intuitively obvious that the optimal coupling π_{opt} between ν_X and $\xi_{(\nu_X,c,0)}$ with $d(x, y) = |y - x|$ moves mass to the left of c further left to ℓ and mass to the right of c further right to L. Any other coupling $\pi \neq \pi_{\text{opt}}$ will map at least some mass $\epsilon > 0$ from the left of c, say from $x \in [\ell, c)$, to L and the same amount of mass from the right of c, say from $x' \in (c, 1]$, to ℓ. Consider now the case $p = 1$: the distance traveled by the ϵ masses increases from $(L - x') + (x - \ell)$ under π_{opt} to $(L - x') + 2(x' - x) + (x - \ell)$ under π, and the cost increases accordingly. We establish this claim more generally for the p-Wasserstein distance in Proposition 1:

Proposition 1. Let ν_X, with center c, be a probability measure defined on $\mathcal{X} = [\ell, L]$, and let $\xi_{(\nu_X,c,\gamma)}$ denote the corresponding maximum separation measure. Then, the optimal p-Wasserstein coupling π^* maps mass $\nu_X[\ell, c) + \gamma\nu_X(c)$ to ℓ, mass $1 - \{\nu_X[\ell, c) + \gamma\nu_X(c)\}$ to L.

Proof. Let X denote the random variable with associated measure ν_X and distribution function F, and let Y denote a random variable with probability measure $\xi_{(\nu_X,c,\gamma)}$ and distribution function G. From Theorem 2.1 in Thorpe (2018), the optimal coupling π_{opt} has c.d.f. $H_{\text{opt}}(x, y) = \min(F(x), G(y))$. We want to show that $H^* = H_{\text{opt}}$ where H^* is the c.d.f. associated with π^*.

To derive H^*, we need only consider $H^*(x, 0)$ and $H^*(x, 1)$ since $\text{supp}(Y) = \{0, 1\}$ and $H^*(x, y) = H^*(x, 0), y \in [0, 1)$:

$$
\begin{aligned}
H^*(x,0) &= \Pr(X \leq x, Y = 0) \\
&= \Pr(Y = 0 \mid X \leq x) \Pr(X \leq x) \\
&= \begin{cases} \Pr(Y = 0 \mid X \leq x) \Pr(X \leq x), & x < c \\ \Pr(Y = 0 \mid X < c) \Pr(X < c) + \Pr(Y = 0 \mid X = c) \Pr(X = c), & x \geq c \end{cases} \\
&= \begin{cases} F(x), & x < c \\ F(c-) + \gamma \nu_X(c), & x \geq c; \end{cases} \\
H^*(x,1) &= \Pr(X \leq x, Y \leq 1) \\
&= \Pr(Y \leq 1 \mid X \leq x) \Pr(X \leq x) \\
&= F(x).
\end{aligned}
$$

Similarly, as $\text{supp}(Y) = \{0, 1\}$, for H_{opt}, we only need to consider $H_{\text{opt}}(x, 0)$ and $H_{\text{opt}}(x, 1)$. Since $G(y) = F(c-) + \gamma \nu_X(c)$ for $y \in [0, 1)$ and $G(y) = 1$ for $y = 1$,

$$
\begin{aligned}
H_{\text{opt}}(x, 0) &= \min(F(x), G(0)) = \min(F(x), F(c-) + \gamma \nu_X(c)) \\
&= \begin{cases} F(x), & x < c \\ F(c-) + \gamma \nu_X(c), & x \geq c \end{cases} \\
H_{\text{opt}}(x, y) &= H_{\text{opt}}(x, 0), \quad \text{if } y \in [0, 1),
\end{aligned}
$$

$$H_{\text{opt}}(x,1) = \min(F(x), G(1)) = F(x). \ \square$$

Substantively, the optimal coupling π_{opt}, by moving people left (right) of c to the left (right) extreme, captures the idea of public opinion change by "movements" of individuals across a "spatial" spectrum of political ideology (Zaller 1992; Downs 1957), inducing a state of polarization as people move "away from the center toward the extremes" (Fiorina and Abrams 2006), with "liberals (conservatives) gravitating more reliably to the liberal (conservative) position" (Zaller 1992 p.102).

Next, we show that the Wasserstein distance satisfies axioms A1-A5.

Proposition 2. $W_p(\nu_X, \xi_{(\nu_X,c,\gamma)})$ satisfies axioms A1-A5.

Proof. Axioms A1 and A5 are satisfied as (2) is a distance between probability distributions. Axiom A2 follows from the definition of the Wasserstein distance with $d(x,y) = |y-x|$: when $X^* = a + bX$ and $Y^* = a + bY$ for $a, b \in \mathbb{R}$, $W_p(\nu_{X^*}, \nu_{Y^*}) = |b| W_p(\nu_X, \nu_Y)$. It remains only to show that Axioms A3 and A4 are satisfied. Because Axiom A2 is satisfied, we proceed, without loss of generality, taking $[\ell, L] = [0,1]$.

To prove Axiom A3 is satisfied, consider first a restricted left spread about c:

$$\begin{aligned}
&W_p^p(\nu_X, \xi_{(\nu_X,c,\gamma)}) - W_p^p(\nu_Y, \xi_{(\nu_X,c,\gamma)}) \\
&\quad = \int_{[0,x_1)} x^p d\nu_X - \int_{[0,x_1)} x^p d\nu_Y + x_1^p(\nu_X(x_1) - \nu_Y(x_1)) + \int_{(x_1,c]} x^p d\nu_X - \int_{(x_1,c]} x^p d\nu_Y \\
&\quad \geq x_1^p[(\nu_X[0,x_1)) - \nu_Y[0,x_1)) + (\nu_X(x_1) - \nu_Y(x_1)) + (\nu_X(x_1,c] - \nu_Y(x_1,c])] \\
&\quad = 0, \qquad (3)
\end{aligned}$$

The case of a restricted right spread about c is proved in an analogous fashion.

That Axiom A4 is satisfied follows from Theorem 3 in Machina and Pratt (1997), which states that the conditions a) ν_Z is obtained from ν_X by a sequence of mean preserving spreads and b) $\int_{[0,1]} u(x)d\nu_Z \leq \int_{[0,1]} u(x)d\nu_X$ for every concave function $u(\cdot)$ are equivalent. For a restricted left mean preserving merge about c

$$W_p^p(\nu_X, \xi_{(\nu_X,c,\gamma)}) - W_p^p(\nu_Y, \xi_{(\nu_X,c,\gamma)}) = \int_{[0,c]} x^p d\nu_X - \int_{[0,c]} x^p d\nu_Y. \tag{4}$$

For $p = 1$, the definition of a mean preserving merge implies (4)$= 0$. For $p > 1$ the result follows from the definition of a restricted left mean preserving merge about c and the convexity of x^p on $[0,1]$. For a restricted right mean preserving merge about c, the result follows similarly from the definition and the convexity of $(1-x)^p$ on $[0,1]$. $\square$

3 A Cross-Country Comparison of Attitudes toward Government Mandates on Vaccination for COVID-19

The COVID-19 pandemic prompted unprecedented governmental actions such as lockdowns, international travel restrictions, and vaccination mandates that stirred much public debate within and across countries. Debates about the safety and efficacy of vaccines, the prioritization of groups during rollout, and, especially, vaccination mandates, were often contentious, and the substance of these disagreements varied across countries. For instance, Democrats and Republicans clashed over vaccination in the US, led by conflicting messages from Biden and Trump (Bolsen and Palm 2022). And in Brazil, where many citizens strongly supported vaccination, followers of (at the time) President Bolsonaro, a COVID-19 denier, displayed sig-

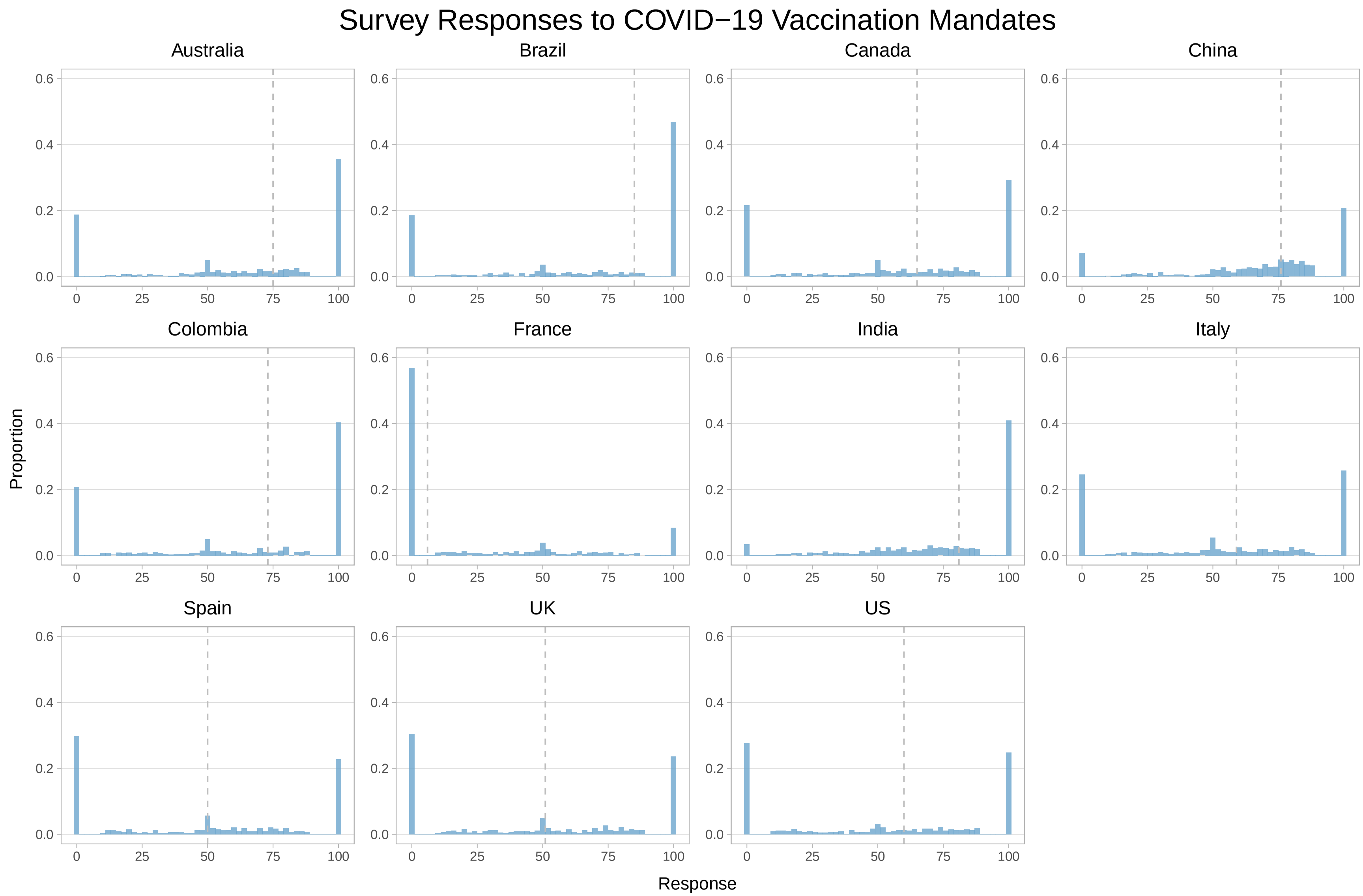

Figure 2: Each histogram features a set of responses from one country to the question "The government should make the COVID-19 vaccine mandatory for everybody," measured on a 0 (disagree) to 100 (agree) VAS scale. Dashed lines indicate the median.

nificantly lower support (Gramacho and Turgeon 2021). On the other hand, Chinese opinions reflected "stabilized public sentiment", possibly due to timely government responses (Han et al. 2020).

To assess the variation in public opinion on vaccine allocation policies, Duch et al. (2021) conducted the Oxford COVID-19 Vaccine Preference and Opinion Survey (CANDOUR), an online survey of 15,536 adults 18 years of age or more, administered between November 2020 and January 2021 in 13 countries, with 1000 to 1500 respondents from each country. In Chile and Uganda, excluded from our analysis, respondents were recruited through Facebook. In the remaining countries (Australia, Brazil, Canada, China, Colombia, France, India, Italy, Spain, the UK, and the US) sampling was conducted by the survey research firm Respondi using respondents from their online panels. Except for Canada and Spain, where demographic and spatial characteristics of sample respondents were close to population totals, post-stratification weights calculated through raking were used to adjust for remaining imbalances between sample and population marginals on region, age, sex and education (Duch et al. 2021; see online appendix).

We apply our index to a question from the survey that asked respondents how much they agree with the following statement: "The government should make the COVID-19 vaccine mandatory for everybody." Responses were recorded as integers on a 0 (disagree) to 100 (agree) visual analogue scale (VAS). Weighted histograms of responses from each country are presented in Figure 2.

3.1 Rescaling $W_p(\nu_X, \xi_{(\nu_X,c,\gamma)})$ to $[0,1]$: The index $I_p(\nu_X, \xi_{(\nu_X,c,\gamma)})$.

To facilitate interpretability, we rescale the p-Wasserstein distance $W_p(\nu_X, \xi_{(\nu_X,c,\gamma)})$ to lie in $[0,1]$ so that the lowest (highest) level of polarization corresponds to a 0 (1) value of the index. To do so we derive the upper and lower bounds of $W_p^p(\nu_X, \xi_{(\nu_X,c,\gamma)})$. The lower bound for $W_p^p(\nu_X, \xi_{(\nu_X,c,\gamma)})$ is obviously 0, which occurs if and only if $\nu_X = \xi_{(\nu_X,c,\gamma)}$, corresponding to an index value $I_p(\nu_X, \xi_{(\nu_X,c,\gamma)}) = 1$. As the upper bound depends on $[\ell, L]$ (see Axiom A2) we assume X has already been transformed to lie in $[0,1]$.

Proposition 3. Let $Q(q) = \inf\{x : \nu_X[0,x] \geq q\}, q \in [0,1]$, denote the quantile function associated with ν_X, and let $\xi_{(\nu_X,c,\gamma)} \equiv \xi_q = q \cdot \delta_0 + (1-q) \cdot \delta_1$, where δ_x is a Dirac measure with mass 1 on x. Then,

$$W_p(\nu_X, \xi_q) \leq W_p(\delta_{Q(q)}, \xi_q) \leq \max(q^{\frac{1}{p}}, (1-q)^{\frac{1}{p}}).$$

Proof:

$$\begin{aligned}
&W_p^p(\nu_X, \xi_q) \\
&\quad = \int_{[0,Q(q))} x^p d\nu_X + \gamma \nu_X(Q) \cdot Q(q)^p + \int_{(Q(q),1]} (1-x)^p d\nu_X + (1-\gamma)\nu_X(Q) \cdot [1-Q(q)]^p \\
&\quad \leq \int_{[0,Q(q))} Q(q)^p d\nu_X + \gamma \nu_X(Q) \cdot Q(q)^p + \int_{(Q(q),1]} [1-Q(q)]^p d\nu_X + (1-\gamma)\nu_X(Q) \cdot [1-Q(q)]^p \\
&\quad = q \cdot [Q(q)]^p + (1-q) \cdot [1-Q(q)]^p = W_p^p(\delta_{Q(q)}, \xi_q) \\
&\quad \leq \max(q, 1-q) = \max[W_p^p(\delta_1, \xi_q), W_p^p(\delta_0, \xi_q)]. \quad \square
\end{aligned}$$

Thus, we define the index $I_p(\nu_X, \xi_{(\nu_X,c,\gamma)}) = 1 - [\max(q^{\frac{1}{p}}, (1-q)^{\frac{1}{p}})]^{-1} W_p(\nu_X, \xi_{(\nu_X,c,\gamma)})$. If $q < .5$ ($q > .5$) the index takes the minimum value of 0 when $W_p^p(\nu_X, \xi_q) = W_p^p(\delta_0, \xi_q) = 1-q$ ($W_p^p(\nu_X, \xi_q) = W_p^p(\delta_1, \xi_q) = q$), i.e., there is a consensus of opinion at 0 (1). Whereas $p' > p \geq 1$ implies $W_{p'} \geq W_p$, this does not hold for the index I_p, due to the rescaling of W_p.

Now, for any value of $q \in [0,1]$, let $Q_p^*(q) = \arg\min_{\{0 \leq Q(q) \leq 1\}} W_p(\delta_{Q(q)}, \xi_q)$ denote the value of $Q(q) \in [0,1]$ that maximizes $I_p(\delta_{Q(q)}, \xi_q)$. It follows from above that $Q_1^*(q) = 1$ if $q < .5$, 0 if $q > .5$ and $\{0,1\}$ if $q = .5$. For $p > 1$, $Q_p^*(q) = \left(\frac{1-q}{q}\right)^{1/(p-1)} / [1 + \left(\frac{1-q}{q}\right)^{1/(p-1)}]$.

Further, for $p \geq 1$, $W_p(\delta_x, \xi_q)$ increases monotonically as $|x - Q_p^*(q)|$ increases. Thus, in contrast to several indices of income bipolarization that take the median $m(\nu_X)$ as the center and result in a value of 0 when the distribution is concentrated on $m(\nu_X)$, no matter where this median is located (Foster and Wolfson 2010; Wang and Tsui 2002), $I_p(\delta_{Q(q)}, \xi_q)$ is sensitive to the location where mass is concentrated, with location at one or both ends of the scale (depending on the choice of q) evidencing less polarization than locations at intermediate points on the scale. These properties of our index allow us to capture important features of opinion polarization that indices insensitive to the location of points where mass is concentrated miss, for example the observation that opinions located at either end of the opinion spectrum are more stable and difficult to change than opinions that are less extreme (Druckman and Leeper 2012; Zaller 1992).

3.2 Results

As previously noted, public opinion researchers typically regard the most polarized state as that in which half of the responses are located at ℓ, the other half at L. This implies a center c equal to the median $m(\nu_X)$, where any choice of c with $P(X \leq m(\nu_X)) \geq 0.5$ and $P(X \geq m(\nu_X)) \leq 0.5$ will result in the same optimal coupling, with half of the responses at ℓ and half at L. After rescaling the responses to lie in the unit interval $[0, 1]$, we define $\xi_{\text{pol}} = 0.5 \cdot \delta_0 + 0.5 \cdot \delta_1$ and estimate the index $I_p(\nu_X, \xi_{\text{pol}}) = 1 - 2^{1/p} \cdot W_p(\nu_X, \xi_{\text{pol}})$. Following Duch et al. (2021), we treat the responses as measurements on a metric scale: if a researcher does not believe a difference of $x_1 - x_2 = \Delta$ units on the scale represents the same amount of opinion as a difference on the scale of $x_3 - x_4 = \Delta$ units, she can consider applying a monotone transformation to the original values so that the transformed values more nearly meet this assumption. Second, the asymptotic 95% confidence intervals we report are obtained under the assumption of simple random sampling (Sommerfeld 2017); to compute these, the R package WPI (Wasserstein Polarization Index) was developed.

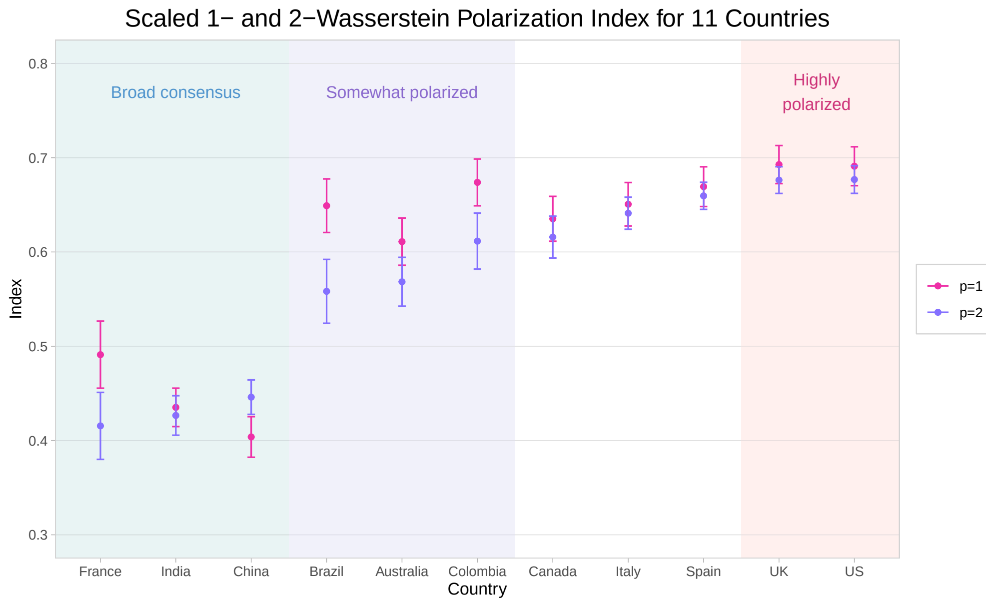

Figure 3: $p = 1$ and $p = 2$ Wasserstein polarization index point estimates of 11 countries in increasing order for $p = 2$. Bars indicate 95% confidence intervals, and the three colored sections correspond to the visual assessment from Duch et al.

In their analysis of attitudes toward mandatory vaccination, Duch et al. (2021) do not define or measure polarization, using visual inspection of the histograms in Figure 2 to classify eight of the 11 countries into three groups: "highly polarized" (US and UK), where the majority of respondents are "either strongly opposed or strongly supportive", "somewhat polarized" (Australia, Brazil, and Colombia), where there is "little middle ground", and "broad consensus" (China, India, and France). Despite the lack of rigor, their grouping is roughly congruent with the ordering of our estimates $I_1(\hat{\nu}_X, \xi_{\text{pol}})$ and $I_2(\hat{\nu}_X, \xi_{\text{pol}})$. However, our estimates are based on an explicit notion of polarization and allow for more refined, quantitative comparisons that take uncertainty into consideration. For instance, within the "somewhat polarized" group, only Australia is significantly less polarized than the "highly polarized" group for $p = 1$. More generally, the eight countries that do not belong to the "broad consensus" group cannot be readily clustered into groups without making somewhat arbitrary distinctions. We also provide point estimates and confidence intervals for Canada, Italy, and Spain, three countries that Duch et al. (2021) do not attempt to classify.

Though our country ordering is essentially the same for both $p = 1$ and $p = 2$, there are several differences, due largely to the fact that I_1 is insensitive to clustering. While China, India and France are the least polarized countries, France appears less polarized than China for $p = 2$, but more polarized for $p = 1$. All else equal, if France and China had equal amounts of clustering around 0 and 0.75, then the distribution clustered around 0.75 would display more polarization for $p = 2$ but not for $p = 1$. Furthermore, France does not exhibit much clustering except near 0, whereas China exhibits stronger clustering on both sides of the median. This also increases polarization for $p = 2$, but not $p = 1$. A similar reversal occurs between Brazil and Australia, but here, for both $p = 1$ and for $p = 2$, the 95% confidence intervals overlap.

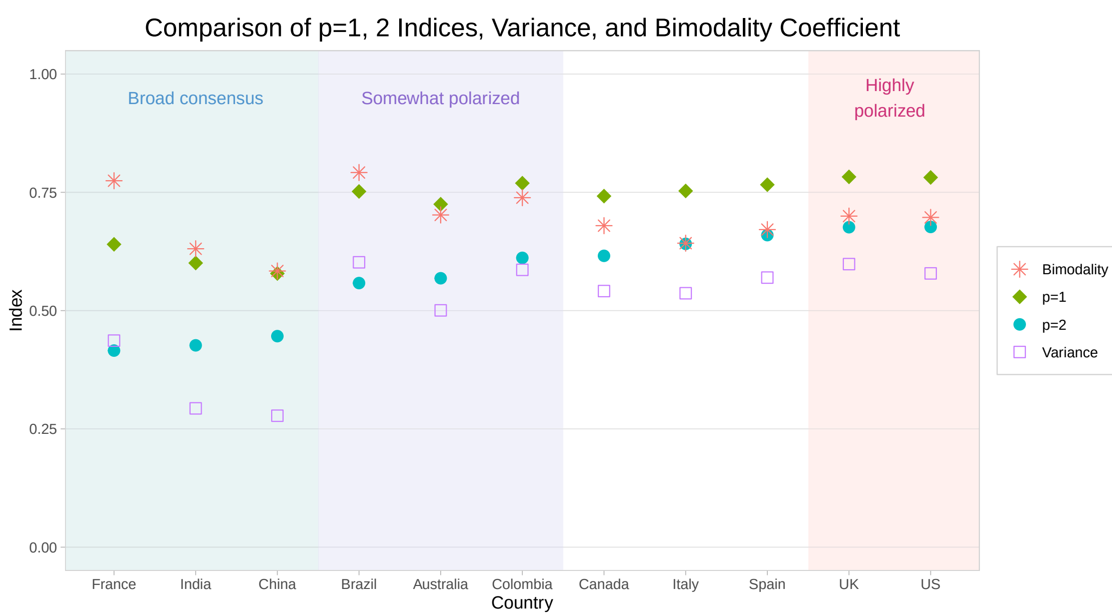

Figure 4: Sarle's sample bimodality coefficient, rescaled variance, and $p = 1$ and $p = 2$ Wasserstein polarization index point estimates of 11 countries in increasing order for $p = 2$. The three colored sections correspond to the visual assessment from Duch et al.

In section 2.2 we showed that several widely used measures of polarization, most notably the variance V and Sarle's sample bimodality coefficient b, do not satisfy all our axioms. Using the estimates in Figure 4, we compare the estimated variance $\hat{V}(\hat{\nu}_X)$ and the bimodality coefficient $b(\hat{\nu}_X)$ with $I_1(\hat{\nu}_X, \xi_{\text{pol}})$ and $I_2(\hat{\nu}_X, \xi_{\text{pol}})$, also with the conclusions reached by Duch et al. Because the variance on $[0, 1]$ takes on a maximum value of .25, to facilitate direct comparison with our indices, we rescaled the variance to take on a maximum value of 1.

The estimated variance $\hat{V}(\hat{\nu}_X)$ tracks the index $I_1(\hat{\nu}_X, \xi_{\text{pol}})$ well for our three least polarized countries. However, by this measure, Brazil and Colombia appear to be more polarized than either the UK or the US, which is not consistent with either our estimates $I_1(\hat{\nu}_X, \xi_{\text{pol}})$ and $I_2(\hat{\nu}_X, \xi_{\text{pol}})$ or the observations in Duch et al. (2021). In addition, $\hat{V}(\hat{\nu}_X)$ is smaller in China than in India, and smaller in India than in France, whereas $I_2(\hat{\nu}_X, \xi_{\text{pol}})$ is larger in China than in India, and larger in India than in France; the reversal between the two orderings is consistent with the fact that all else equal, mean preserving clustering decreases $\hat{V}(\hat{\nu}_X)$ and increases $I_2(\hat{\nu}_X, \xi_{\text{pol}})$,

Sarle's sample bimodality coefficient already lies in $[0, 1]$ and does not require rescaling. Figure 4 demonstrates that the coefficient largely does not track the other measures. For instance, France has the second highest value of $b(\hat{\nu}_X)$, but every other measure suggests France is one of the least polarized nations. In addition, by this measure, France and the three "somewhat polarized" countries appear to be more polarized than the "highly polarized" countries identified by our measures and by Duch et al. Further, as is apparent from inspection of the histograms in Figure 2, whereas the distributions in Italy, the U.S, the U.K, and Spain appear to exhibit some bimodality, the three countries with the highest bimodality coefficients (Brazil, France, and Colombia) exhibit clustering at one of the poles, but little clustering at other parts of the distribution. It seems reasonable to conclude that

b is not measuring bimodality here.

4 Discussion

To study bipolarization, public opinion researchers use ad hoc measures that fail to align with their notions of this concept. This paper sets forth key properties an index of bipolarization should satisy and proposes a p-Wasserstein index consistent with these properties. To illustrate the index, we re-analyze an item from a study by Duch et al. (2021) that asked respondents in a number of countries if they thought vaccination against COVID-19 should be mandatory in their country. Among the 11 countries considered, China, India and France are the least polarized, while the UK and US are the most polarized.

Several extensions of our approach would be useful. First, we compared interval estimates of the p-Wasserstein index in different countries. More common are studies that track the distribution of an item or items over time within a country to assess whether or not polarization has increased during the period under investigation. (DiMaggio, Evans, and Bryson 1996; Fiorina and Abrams 2008; Mouw and Sobel 2001; Lelkes 2016). While the p-Wasserstein index can also be applied directly in this context, it would also be useful to develop frameworks for modeling temporal trends in the index and for testing hypotheses about differences in index values between groups and/or over time.

Second, many survey items are measured on ordinal scales with 7 or fewer categories, and most researchers treat responses to such scales as cardinal. While it is trivial to extend our index to the ordinal case by using a suitable cost metric to measure the difficulty of moving between different categories, this begs the question of how to choose such a metric. We are currently studying several approaches: clearly, this requires incorporating additional information and assumptions.

Third, public opinion researchers often want to aggregate the results from different opinion items to form a summary judgement. To study ideological divergence, some researchers do not attempt to mathematically combine the responses from different items into a composite measure: for example, Fiorina (2017, p. 28) argued, on the basis of visual inspection of graphs of five NES items asked of respondents in 1984 and 2016, that "the distributions in 2016 maintain the same generally centrist shape as they did in 1984." In contrast, Hill and Tausanovitch (2015) used an item response model with 67 NES items to estimate respondent's positions on a unidimensional "ideological" scale. To assess whether or not polarization had increased over time, they compared the variance of the distributions of positions at the different time points. In future work, we intend to extend our approach to develop a principled polarization index based on multiple items.

We also believe the Wasserstein distance can be used to construct better measures of other types of polarization recognized in the literature. More recently, political scientists and others have called attention to increasing conflict among partisans of different parties, and the quantification of partisan polarization has taken on great importance. Here, instead of measuring polarization among the public at large by comparing the observed distribution of responses on an item to a fixed distribution with all probability concentrated on the left and role poles of the scale, as we have done, the aim is to compare the distribution of the item among Republicans with the distribution of the item among Democrats. To measure the so-called "partisan gap", researchers often use the between group difference in means or median on the item, using trends in these quantities to ascertain whether or not "partisan polarization" has increased, decreased, or neither (Hill and Tausanovitch 2015, Enders 2021). But it is easy to construct examples where such measures may be misleading: consider the case of a 0-100 interval scale where 50% of Republicans are located at 25 and 50% are located at 75, and all Democrats at 50, in which case both medians and means of each of the distributions are located at 50, leading to the conclusion that there is no "partisan gap". We believe the Wasserstein distance between the two distributions could be used to better inform the issue of partisan polarization than the possibly misleading comparisons based on means and/or medians.

To the best of our knowledge, ours is the first paper to use the theory of optimal transport to develop a sociopolitical index. Whereas our index is defined as a comparison of a distribution ν_X to a prescribed distribution with mass concentrated on the minimum and maximum scale values, the Wasserstein distance is frequently used to compare arbitrary distributions ν_X and ν_Y, and we advocated using it to compare the distribution of Democrats and Republicans on public opinion items. More generally in social research, where comparisons between two groups are ubiquitous (e.g., educational attainments of Blacks and whites or males and females) the disparity between groups is often operationalized as a difference in means or medians of an outcome of interest; the Wasserstein distance, with its capacity to capture the relative geometry between distributions, is able to detect dissimilarities between distributions that are missed by these cruder summaries. Thus, we believe the Wasserstein distance has a potentially important role to play in social research for conceptualizing, measuring and testing between group similarities and dissimilarities.

Acknowledgements

For helpful comments on a previous draft, we thank David Banks, Naoki Egami, Bodhi Sen, Andrés Riveros Valdevenito, Zhiliang Ying, Jacob Montgomery and panel participants at

the 41st Annual Meeting of the Society for Political Methodology.

Appendix: Computation and Inference

The p-Wasserstein distance between discrete measures ν_X and ν_Y on a finite support $\mathcal{X}$ can be expressed as:

$$[W_p(\nu_X, \nu_Y)]^p = \inf_{\pi \in \Pi(\nu_X, \nu_Y)} \sum_{x,y \in \mathcal{X}} d^p(x,y)\pi(x,y),$$

where $\Pi(\nu_X, \nu_Y)$ is the set of all joint probability measures with marginals ν_X and ν_Y. Let $\{x_i\}_{i=1}^M$ denote the support of ν_X and $\{y_j\}_{j=1}^N$ the support of ν_Y. Expressing the two measures as "superpositions" $\nu_X = \sum_{i=1}^M p_i \delta_{x_i}$ and $\nu_Y = \sum_{j=1}^N q_j \delta_{y_j}$ (Solomon 2018), it is evident that the p-Wasserstein distance is the solution to the linear program

$$\begin{aligned} \min_{\boldsymbol{\pi}} \quad & \sum_{i,j} d^p(x_i, y_j)\pi_{ij} \\ \text{s.t.} \quad & \sum_i \pi_{ij} = q_j, \sum_j \pi_{ij} = p_i, \pi_{ij} \geq 0. \end{aligned}$$

The dual program is given by:

$$\begin{aligned} \max_{(\mathbf{v},\mathbf{w}) \in \mathbb{R}^M \times \mathbb{R}^N} \quad & \sum_{i=1}^M v_i p_i + \sum_{j=1}^N w_j q_j \\ \text{s.t.} \quad & v_i + w_j \leq d^p(x_i, y_j). \end{aligned}$$

with dual variables $\mathbf{v}, \mathbf{w}$.

Using the dual, Sommerfeld (2017) derives the asymptotic distribution of the empirical Wasserstein distance between $\hat{\nu}_n$, the empirical measure of ν based on n observations, and a known measure ξ (Theorem 8(a)), from which asymptotic confidence intervals follow:

Assume ν, ξ on $\mathcal{X}$ satisfy the non-degeneracy condition: $\sum_{x \in A} \nu(x) \neq \sum_{y \in B} \xi(y)$ for all proper subsets A, B of $\mathcal{X}$. Let $\mathbf{v}^*, \mathbf{w}^*$ denote a solution to the dual problem of $W_p(\hat{\nu}_n, \xi)$. Then, an asymptotic $(1-\alpha) \times 100\%$ confidence interval for $W_p(\hat{\nu}_n, \xi)$ is given by

$$\left[W_p(\hat{\nu}_n, \xi) - \frac{z_{\alpha/2}}{\sqrt{n}p} W_p^{1-p}(\hat{\nu}_n, \xi)\sigma(\hat{\nu}_n, \xi), W_p(\hat{\nu}_n, \xi) + \frac{z_{\alpha/2}}{\sqrt{n}p} W_p^{1-p}(\hat{\nu}_n, \xi)\sigma(\hat{\nu}_n, \xi)\right],$$

where $\sigma^2(\hat{\nu}_n, \xi) = \sum_{x \in \mathcal{X}} \mathbf{v}^*(x)^2 \hat{\nu}_n(x) - \left(\sum_{x \in \mathcal{X}} \mathbf{v}^*(x)\hat{\nu}_n(x)\right)^2$and $z_{\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution. Thus, for our rescaled index $I_p(\hat{\nu}_n, \xi) = 1 - \max(q^{1/p}, (1-q)^{1/p})^{-1} \cdot W_p(\hat{\nu}_n, \xi)$, an asymptotic $(1-\alpha) \times 100\%$ confidence interval is

$$\Bigg[1 - \max(q^{1/p}, (1-q)^{1/p})\Big\{W_p(\hat{\nu}_n, \xi) + \frac{z_{\alpha/2}}{\sqrt{n}p} W_p^{1-p}(\hat{\nu}_n, \xi)\sigma(\hat{\nu}_n, \xi)\Big\},$$
$$1 - \max(q^{1/p}, (1-q)^{1/p})\Big\{W_p(\hat{\nu}_n, \xi) - \frac{z_{\alpha/2}}{\sqrt{n}p} W_p^{1-p}(\hat{\nu}_n, \xi)\sigma(\hat{\nu}_n, \xi)\Big\}\Bigg].$$

The code used to compute the confidence intervals in the empirical application is given in the R package "WPI": Wasserstein Polarization Index.

References

Abramowitz, A. I. and Saunders, K. L. (2008). "Is Polarization a Myth?" *The Journal of Politics* 70(2), pp. 542–555.

Allison, R. and Foster, J. E. (2004). "Measuring Health Inequality Using Qualitative Data". *Journal of Health Economics* 23(3), pp. 505–524.

Anderson, G. (2004). “Making Inferences About the Polarization, Welfare and Poverty of Nations: A Study of 101 Countries 1970–1995”. *Journal of Applied Econometrics* 19(5), pp. 537–550.

Apouey, B. (2007). “Measuring Health Polarization With Self-assessed Health Data”. *Health Economics* 16(9), pp. 875–894.

Apouey, B. and Silber, J. (2013). “Inequality and Bi-polarization in Socioeconomic Status and Health: Ordinal Approaches”. *Health and inequality*. Vol. 21. Emerald Group Publishing Limited, pp. 77–109.

Baldassarri, D. and Gelman, A. (2008). “Partisans without Constraint: Political Polarization and Trends in American Public Opinion”. *American Journal of Sociology* 114(2), pp. 408–446.

Bauer, P. (2019). *Conceptualizing and Measuring Polarization: A review.*

Bauer, P. C., Barberá, P., Ackermann, K., and Venetz, A. (2017). “Is the Left-Right Scale a Valid Measure of Ideology?: Individual-Level Variation in Associations with “Left” and “Right” and Left-Right Self-Placement”. *Political Behavior* 39(3), pp. 553–583.

Bolsen, T. and Palm, R. (2022). “Politicization and COVID-19 Vaccine Resistance in the U.S.” *Molecular Biology and Clinical Medicine in the Age of Politicization*. Ed. by T. Bolsen and R. Palm. Vol. 188. Progress in Molecular Biology and Translational Science 1. Academic Press, pp. 81–100.

DiMaggio, P., Evans, J., and Bryson, B. (1996). “Have American’s Social Attitudes Become More Polarized?” *American Journal of Sociology* 102(3), pp. 690–755.

Downs, A. (1957). *An Economic Theory of Democracy*. Harper & Brothers.

Druckman, J. N. and Leeper, T. J. (2012). “Is Public Opinion Stable? Resolving the Micro/Macro Disconnect in Studies of Public Opinion”. *Daedalus* 141(4), pp. 50–68.

Druckman, J. N. and Levy, J. (2022). "Chapter 18: Affective Polarization in the American Public". *Handbook on Politics and Public Opinion*. Cheltenham, UK: Edward Elgar Publishing, pp. 257–270.

Duch, R., Roope, L. S. J., Violato, M., Becerra, M. F., Robinson, T. S., Bonnefon, J.-F., Friedman, J., Loewen, P. J., Mamidi, P., Melegaro, A., Blanco, M., Vargas, J., Seither, J., Candio, P., Cruz, A. G., Hua, X., Barnett, A., and Clarke, P. M. (2021). "Citizens From 13 Countries Share Similar Preferences for COVID-19 Vaccine Allocation Priorities". *Proceedings of the National Academy of Sciences* 118(38), e2026382118.

Duclos, J.-Y., Esteban, J., and Ray, D. (2004). "Polarization: Concepts, Measurement, Estimation". *Econometrica* 72(6), pp. 1737–1772.

Enders, A. M. (2021). "Issues Versus Affect: How Do Elite and Mass Polarization Compare?" *The Journal of Politics* 83(4), pp. 1872–1877.

Esteban, J., Gradín, C., and Ray, D. (2007). "An Extension of a Measure of Polarization, With an Application to the Income Distribution of Five OECD Countries". *The Journal of Economic Inequality* 5(1), pp. 1–19.

Esteban, J. and Ray, D. (1994). "On the Measurement of Polarization". *Econometrica* 62(4), pp. 819–851.

— (2012). "Comparing Polarization Measures". *The Oxford Handbook of the Economics of Peace and Conflict*.

Fiorina, M. P. (2017). *Unstable Majorities: Polarization, Party Sorting, and Political Stalemate*. Hoover Institution Press.

Fiorina, M. P. and Abrams, S. J. (2008). "Political Polarization in the American Public". *Annual Review of Political Science* 11(1), pp. 563–588.

Fiorina, M. P. and Levendusky, M. S. (2006). “Disconnected: The political class versus the people”. *Red and Blue Nation?: Characteristics and Causes of America’s Polarized Politics*. Brookings Institution Press, pp. 49–71.

Foster, J. and Wolfson, M. (2010). “Polarization and the Decline of the Middle Class in Canada and the USA”. *Journal of Economic Inequality* 8, pp. 247–273.

Gilens, M. and Page, B. I. (2014). “Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens”. *Perspectives on Politics* 12(3), pp. 564–581.

Gramacho, W. G. and Turgeon, M. (2021). “When Politics Collides With Public Health: COVID-19 Vaccine Country of Origin and Vaccination Acceptance in Brazil”. *Vaccine* 39(19), pp. 2608–2612.

Han, X., Wang, J., Zhang, M., and Wang, X. (2020). “Using Social Media to Mine and Analyze Public Opinion Related to COVID-19 in China”. *International journal of environmental research and public health* 17(8), p. 2788.

Hill, S. J. and Tausanovitch, C. (2015). “A Disconnect in Representation? Comparison of Trends in Congressional and Public Polarization”. *The Journal of Politics* 77(4), pp. 1058–1075.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., and Westwood, S. J. (2019). “The Origins and Consequences of Affective Polarization in the United States”. *Annual Review of Political Science* 22(1), pp. 129–146.

Kim, H. and Fording, R. C. (2003). “Voter Ideology in Western Democracies: An Update”. *European Journal of Political Research* 42(1), pp. 95–105.

Kobus, M. (2015). “Polarization Measurement for Ordinal Data”. *The Journal of Economic Inequality* 13(2), pp. 275–297.

Koudenburg, N., Kiers, H. A. L., and Kashima, Y. (2021). "A New Opinion Polarization Index Developed by Integrating Expert Judgments". *Frontiers in Psychology* 12.

Kozlowski, A. C. and Murphy, J. P. (2021). "Issue Alignment and Partisanship in the American Public: Revisiting the 'Partisans Without Constraint' Thesis". *Social Science Research* 94, p. 102498.

Lelkes, Y. (2016). "Mass Polarization: Manifestations and Measurements". *Public Opinion Quarterly* 80(S1), pp. 392–410.

Levendusky, M. (2009). *The Partisan Sort: How Liberals Became Democrats and Conservatives Became Republicans.* Chicago Studies in American Politics. University of Chicago Press.

Machina, M. J. and Pratt, J. W. (1997). "Increasing Risk: Some Direct Constructions". *Journal of Risk and Uncertainty* 14, pp. 103–127.

Melki, M. and Pickering, A. (2014). "Ideological Polarization and the Media". *Economics Letters* 125(1), pp. 36–39.

Merkley, E. (2022). "Polarization Eh? Ideological Divergence and Partisan Sorting in the Canadian Mass Public". *Public Opinion Quarterly* 86(4), pp. 932–943.

Mouw, T. and Sobel, M. E. (2001). "Culture Wars and Opinion Polarization: The Case of Abortion". *American Journal of Sociology* 106(4), pp. 913–943.

Peyré, G. and Cuturi, M. (2019). "Computational Optimal Transport: With Applications to Data Science". *Foundations and Trends in Machine Learning* 11(5-6), pp. 355–607.

Poole, K. T. and Rosenthal, H. L. (2011). *Ideology and Congress.* Transaction Publishers.

Rothschild, M. and Stiglitz, J. E. (1970). "Increasing Risk: I. A Definition". *Journal of Economic Theory* 2(3), pp. 225–243.

Sarkar, S. and Santra, S. (2020). “Extending the Approaches to Polarization Ordering of Ordinal Variables”. *The Journal of Economic Inequality* 18(3), pp. 421–440.

Simas, E. N., Clifford, S., and Kirkland, J. H. (2020). “How Empathic Concern Fuels Political Polarization”. *American Political Science Review* 114(1), pp. 258–269.

Solomon, J. (2018). *Optimal Transport on Discrete Domains.*

Sommerfeld, M. (2017). “Wasserstein Distance on Finite Spaces: Statistical Inference and Algorithms”. PhD thesis. Niedersächsische Staats-und Universitätsbibliothek Göttingen.

Strijbis, O., Helmer, J., and De Wilde, P. (2020). “A Cosmopolitan–Communitarian Cleavage Around the World? Evidence From Ideological Polarization and Party–Voter Linkages.” *Acta Politica* 55(3), pp. 408–431.

Thorpe, M. (2018). “Introduction to Optimal Transport”. *Notes of Course at University of Cambridge.*

Villani, C. (2009). *Optimal Transport: Old and New.* Vol. 338. Springer.

Wang, Y.-Q. and Tsui, K.-Y. (2000). “Polarization Orderings and New Classes of Polarization Indices”. *Journal of Public Economic Theory* 2(3), pp. 349–363.

Zaller, J. R. (1992). *The Nature and Origins of Mass Opinion.* Cambridge Studies in Public Opinion and Political Psychology. Cambridge University Press.