

BACKGROUND AND QUALIFICATIONS

Education

3. I received my B.A. from Grinnell College (with honors, Phi Beta Kappa) and my Ph.D. in Mathematics from the University of Pennsylvania.

Academic Employment

4. I currently am a Full Professor and the Chair of the Mathematics Department at Saint Joseph's University. Prior to that, I taught mathematics at Haverford College, SUNY Stony Brook, Bryn Mawr College, Williams College, the University of Pennsylvania, and Suffolk University. My research was supported by National Science Foundation grants during portions of my time at several of these institutions.

Publications and Presentations

5. I am the author of three textbooks. I have published 24 research papers in refereed journals, including some of the top journals in mathematics. I have presented research talks at about 50 conferences and seminars including in Germany, Argentina, Mexico, Canada, and Spain.

6. For the last four years, my primary research focus has been mathematical questions related to redistricting.

7. My first publication on this topic was *Measuring Political Gerrymandering*, which was published in *American Mathematical Monthly*, the most widely read mathematics journal in the world. This peer-reviewed paper explored the mathematical properties of the *efficiency gap* and its variants. I was invited to speak about the results of this peer-reviewed paper at the *Geometry of Redistricting Workshop* in San Francisco and also at the *University of Arizona Conference on Redistricting*.

8. My second research paper on this topic was *Clustering and Expected Seat-Share for District Maps*, published in *Communications in Statistics – Theory and Methods*. In this peer-reviewed paper, I used ensemble methods to quantify and precisely explore how a political party's political geography (particularly its clustering) affects the outcome of an election. This peer-reviewed paper included an ensemble analysis for congressional maps of Pennsylvania.

9. My third research paper was *Spanning Tree Bounds for Grid Graphs*. All of the popular methods for ensemble analysis of district maps today rely on spanning trees. There are several purely mathematical questions about spanning trees that the mathematics community hopes to solve in order to improve the theoretical foundation on which these methods are built. In this peer-reviewed paper, I addressed and solved a specific mathematical question about counting spanning trees, providing a better understanding of why today's ensemble methods naturally generate compact districts without needing to be specifically directed to do so.

10. In short, my recent research has focused on the mathematics that undergirds modern algorithms for analyzing district maps.

11. A copy of my curriculum vitae is attached as Exhibit A to this affidavit.

SCOPE OF ENGAGEMENT AND COMPENSATION

12. I have been retained by Cuti Hecker Wang LLP, counsel for Respondent Senate Majority Leader and President Pro Tempore of the Senate Andrea Stewart-Cousins, and asked to opine on the validity of the analysis used and the conclusions drawn in the report submitted by Sean Trende.

13. I am being compensated at a rate of \$400.00 per hour. My compensation does not depend in any way on the outcome of the case or on the opinions or testimony that I provide.

MATERIALS REVIEWED

14. In connection with preparing this testimony and providing the opinions expressed herein, I have reviewed the following materials:

- Report of Sean Trende submitted on behalf of the Petitioners in this case;
- Relevant portions of Article III, Section 4(c) of the New York Constitution setting forth applicable redistricting criteria; and
- McCartan & Imai, *Sequential Monte Carlo for Sampling Balanced and Compact Redistricting Plans*.

SUMMARY OF EXPERT OPINIONS

15. Mr. Trende's stated opinion is that the enacted congressional map was drawn for the purpose of favoring the Democratic Party. Based on my analysis of the report and its methodology, I hold the following opinions to a high degree of professional certainty:

- a. Mr. Trende's data clearly supports the opposite conclusion: that the enacted congressional map favors the Republican Party. For example, his data shows that the enacted map gives the Democrats 22 seats, whereas almost every one of the 5,000 randomly generated maps in his ensemble gives the Democrats at least 22 seats, substantially more than half of those maps give the Democrats at least 23 seats, and a good number of the maps give the Democrats 24 or 25 seats. If we take at face value Mr. Trende's assertion that his 5,000 maps represent "what maps would tend to look like in New York if they were drawn without respect for politics," then the only reasonable conclusion is that the enacted map is significantly Republican-favoring relative to maps drawn without attention to politics.

- b. Mr. Trende’s methodology is so deeply flawed that the ensemble he created is not a representative sample of maps that could be drawn without partisan considerations, and the results he produced have no meaningful statistical value. Among other significant flaws in Mr. Trende’s methodology, his model fails to account for a number of the redistricting criteria that are required by New York law. Accordingly, Mr. Trende’s analysis is unreliable.

Background on Methods Used by Mr. Trende

Background on Ensemble Methods

16. Ensemble methods for assessing district maps sprung up in the previous decade and quickly became a mainstay of partisan gerrymandering litigation. The basic idea is to construct an *ensemble* of thousands or millions of randomly generated district maps. The enacted map is then compared to the ensemble. If the enacted map is a statistical outlier – for example, if the enacted map is likely to elect significantly more members from one political party than do the maps in the ensemble – this is taken as evidence that the enacted map was drawn with partisan intent.

17. For such a conclusion to be reasonable, the algorithm generating the ensemble of random maps cannot use any partisan data. But a partisan-blind algorithm is not enough. To yield meaningful and reliable statistical conclusions, it is imperative that the algorithm sample from a prescribed *target distribution*. To understand what is meant by a “target distribution,” imagine a jurisdiction with such a small population that there are only 100 possible ways to apportion the population into districts. If one wanted to build an ensemble of 10 such maps, one could direct a computer to do the equivalent of putting the 100 maps into a hat and selecting 10 of them at random with equal likeliness. That is called a *uniform target distribution* because all

of the compliant maps are equally likely to appear in the ensemble. One also could direct a computer to weight the 100 maps such that the ones with better compactness scores are more likely to be selected for the ensemble. Such a *weighted target distribution* might be reasonable, provided that the weighting formula is appropriate. But it would be problematic, and would fatally undermine the results, if some of the maps generated by the algorithm failed to comply with applicable redistricting requirements.

18. In real states, there are countless numbers of potential compliant maps, so a computer cannot do the equivalent of identifying all of them, putting them in a hat, and selecting an ensemble from among the maps in the hat. Mathematicians therefore have developed methods to sample from the uniform target distribution (or any prescribed target distribution) even when the complete sets involved would be too large to store in a computer. These sampling methods are often called *Monte Carlo* methods. They are at the heart of most of the commonly used algorithms for generating ensembles of district plans.

Background on the Gerrymandering Index

19. The partisan analysis in Mr. Trende's report is based almost exclusively on the *gerrymandering index*. This measurement was first proposed by a research team based at Duke University; see *Redistricting: Drawing the Line*, Bangia, et al., 2017, arXiv:1704.03360. Mr. Trende's report appears to mischaracterize what the gerrymandering index does and does not measure. It therefore is necessary to explain the gerrymandering index and the *ordered district plots* on which this index is based.

20. *Ordered district plots* and the *gerrymandering index* are best understood with a hypothetical example of a state with three congressional districts. The outcome of an election using the enacted map can be represented by a vector, which means a list of three numbers –

namely, the Democratic vote shares in the three districts listed in increasing order. For example, suppose that

$$\textit{Enacted} = (.25, .60, .75)$$

which means that in the most Republican-leaning district, 25% of the votes (only counting votes for the two major parties) went to the Democratic candidate; in the middle district, 60% went to the Democratic candidate, and in the most Democrat-leaning district, 75% went to the Democratic candidate. This vector is represented by the three black dots in the ordered district plot shown in Figure 1 below. Notice that two of the three seats went to the Democratic candidate (because two of black dots are above the dashed .5 line).

21. Imagine a hypothetical ensemble of five randomly generated maps. The simulated election outcome in each map yields a corresponding vector that is plotted as three green dots in Figure 1 below. In this hypothetical example,

$$\textit{Ensemble Average} = (.19, .70, .72)$$

which means that, among the five maps in the ensemble, the Democratic vote shares in their most Republican-leaning districts averaged to 19%, the Democratic vote shares in their middle districts averaged to 70%, and the Democratic vote shares in their most Democrat-leaning districts averaged to 72%. In other words, the average heights of the three columns of green dots are respectively .19, .70, and .72.

22. The gerrymandering index is just the distance between the Enacted vector and the Ensemble Average vector (considered as points in three-dimensional space); in other words:

$$\textit{gerrymandering index} = \sqrt{(.25 - .19)^2 + (.60 - .70)^2 + (.75 - .72)^2} \approx .12$$

23. Notice that the gerrymandering index is larger when the three illustrated blue brackets are taller:

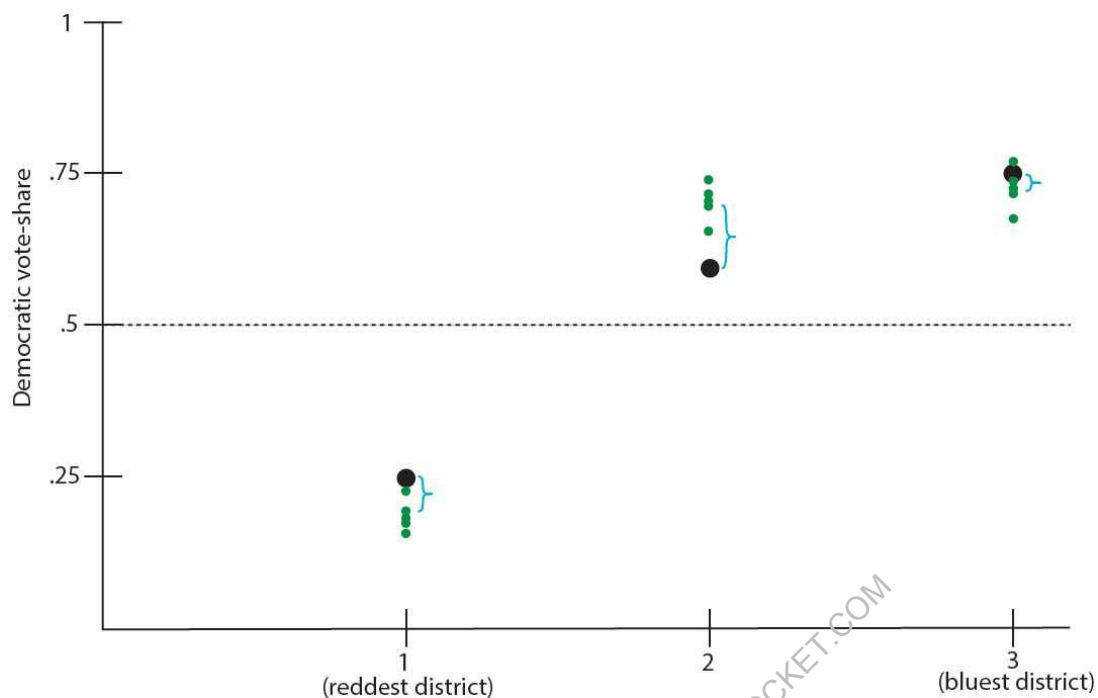


Figure 1: The ordered district plot for a hypothetical toy example

24. Because Mr. Trende’s report includes ordered district plots, it is important to note the types of information that can be recovered from such a plot. In our hypothetical example, we know from the plot that all five maps in the ensemble elected two Democrats and one Republican (because all of the green dots at horizontal position “1” are below the 50% line, while all of the green dots at positions “2” and “3” are above it).

25. Thus, the gerrymandering index summarizes how closely the enacted map’s vector matches the vectors of the maps in the ensemble. Importantly, there are two things that the gerrymandering index does *not* do:

- The gerrymandering index does *not* provide any information about *which party* is favored by the enacted map relative to the ensemble, or even whether there is a favored party.
- The gerrymandering index does *not* provide any information about whether the enacted map discourages competitive districts relative to the ensemble.

26. There are a variety of factors that can cause the gerrymandering index to be large. This can happen if the enacted map is systematically biased towards the Republican party (relative

to the ensemble). It can happen if the enacted map is systematically biased towards the Democratic party. It can happen if the enacted map is drawn to protect incumbents from both parties without any overall partisan lean. It can happen if the enacted map has fewer competitive districts than average maps. It can happen if the enacted map has more competitive districts than average maps. It can happen because of differences between the enacted map and the ensemble that do not influence the election outcome at all. It can happen for any mixture of these reasons.

ANALYSIS OF MR. TRENDE'S CONCLUSIONS

27. As detailed below, Mr. Trende's methodology has such substantial flaws that his model has little if any statistical value. Notwithstanding those methodological flaws, I begin by taking at face value Mr. Trende's claim that his ensemble of 5,000 maps represents "what maps would tend to look like in New York if they were drawn without respect for politics."

Conclusions about Partisan Bias

28. There are standard methods for measuring the partisan bias of an enacted map relative to an ensemble. The simplest method is to plot a histogram showing how many districts are likely to be won by Democrats (among all of the maps in the ensemble) and compare this to the number of districts that are likely to be won by Democrats using the enacted map. If the enacted map deviates too far from the average, this is seen as evidence that the map was drawn to favor one political party.

29. Mr. Trende did not perform this type of analysis. But the information found in his ordered district plots (pages 15 and 21) can be used to determine what such a histogram and analysis would have looked like. Under such an analysis, it is 100% certain that the congressional map has a Republican-favoring partisan bias relative to Mr. Trende's ensemble. This is the opposite of what he claims.

30. The chart on page 15 of Mr. Trende's report, titled "Democratic Vote Share by Simulated District," indicates that the enacted map is likely to give the Democrats 22 seats. (This is reflected by the black dots for the districts numbered 1 through 4 below the dashed 50% line, indicating higher Republican vote shares, and the black dots for the districts numbered 5 through 26 above the 50% line, indicating higher Democratic vote shares.) By comparison, virtually all of the 5,000 maps in Mr. Trende's ensemble are likely to give the Democrats at least 22 seats; substantially more than half are likely to give the Democrats at least 23 seats; and a good number are likely to give the Democrats 24 or 25 seats. (This is reflected by which proportion of the shaded coloring is red and below the 50% line, or blue and above the 50% line, for each of the numbered districts.) On average, the maps in his ensemble clearly are likely to give the Democrats more seats than does the enacted map.

31. Indeed, a careful review of the chart on page 15 of Mr. Trende's report shows that the outcome that is overwhelmingly most likely statistically is a plan that gives the Democrats 23 seats. The next most likely outcome is not a plan that gives Democrats 22 seats, like the enacted plan, but rather a plan that gives Democrats 24 seats, *i.e.*, 2 more seats than the enacted plan.

32. In other words, the only conclusion that can be drawn from Mr. Trende's data is that the enacted map is significantly Republican-favoring relative to the maps in Mr. Trende's ensemble.

33. Mr. Trende suggests incorrectly that the gerrymandering index proves partisan bias. On page 15 of his report, he writes that "[i]f the Enacted Congressional Map was not drawn to favor or disfavor a political party, or did so only moderately, it should hew close to the results produced by simulated maps [equivalently, it should have a small gerrymandering index]." That is wrong. Partisan lean is only one of many factors that can make the

gerrymandering index high, and to the extent that partisan lean contributed to the high gerrymandering index in Mr. Trende's analysis, it was clearly a *Republican*-favoring lean that made the gerrymandering index high.

34. In summary, Mr. Trende's data proves the opposite of what he claims. The enacted congressional map is substantially more favorable to Republicans than the maps in his ensemble.

ANALYSIS OF MR. TRENDE'S METHODOLOGY

35. In this section, I will analyze the methodology that Mr. Trende used to construct his ensemble, and I will examine whether his ensemble is a representative sample of nonpartisan maps.

Mr. Trende's Lack of Relevant Experience

36. Mr. Trende notes in his expert report that he is currently a doctoral candidate and has not yet received his Ph.D.

37. It is my understanding that Mr. Trende has never published a peer-reviewed article, much less a peer-reviewed article concerning ensemble analysis or any other mathematical analysis of gerrymandering.

38. Mr. Trende's report does not indicate a very deep understanding of the underlying algorithm he employs. For example, he suggests on page 8 that spanning trees are constructed by breaking adjacencies (which is not what the algorithm he employs does because this would not work). On the same page, he incorrectly asserts that sets with more than two precincts are guaranteed to have multiple spanning trees. On page 7, he describes ensemble algorithms as "potentially subject to a variety of parameters," which suggests to me a surface-level engagement with the interface of an algorithm that someone else coded.

Lack of Reproducibility

39. Mr. Trende's report contains almost no information about his methodology, which makes it impossible for me to reproduce his result or definitively diagnose his errors. His report lacks the key information that would be required to check the validity of his ensemble. This omission alone is cause for serious concern.

40. What I do know based on the limited information that Mr. Trende provides in his report is that Mr. Trende uses the *redist* package in the *R* programming language (available at <https://alarm-redist.github.io/posts/2021-04-02-redist-300/>). This package includes coding for several popular algorithms to generate ensembles. It appears from Mr. Trende's report that he used the *redist_smc* package, which is based on the very new *Sequential Monte Carlo* technique developed in the paper: McCartan & Imai, *Sequential Monte Carlo for Sampling Balanced and Compact Redistricting Plans*.

Missing Constitutional Requirements

41. One flaw in Mr. Trende's methodology is that he only incorporates a subset of the criteria that the New York Constitution requires to be taken into consideration in redistricting.

42. Mr. Trende's report indicates that he instructed the simulation model to consider three criteria in generating the simulated New York districts: (1) districts must be equipopulous within a population tolerance of +/- 1%; (2) districts must be reasonably compact; and (3) districts should avoid county splits. (His underlying algorithm also guarantees that districts are contiguous.)

43. Mr. Trende's model does not in any way take into account a number of the criteria that the New York Constitution states shall be used in redistricting. Among other factors, Mr. Trende's model does not take into account the following considerations:

- whether the districts would result in the denial or abridgement of racial or language minority voting rights;
- whether the districts are drawn so that racial or minority language groups do not have less opportunity to participate in the political process than other members of the electorate and to elect representatives of their choice;
- the maintenance of cores of existing districts;
- the maintenance of cities;
- the maintenance of towns; or
- the maintenance of communities of interest.

44. Mr. Trende does not explain why, or even acknowledge that, his model fails to incorporate those constitutionally-required considerations.

45. Because Mr. Trende's model does not include those considerations, the model is incapable of doing what Mr. Trende contends that it does, *i.e.*, "create an 'ensemble' of maps that reflect what we would expect in a state if maps were drawn without respect to partisan criteria." In Mr. Trende's model, "partisan criteria" is not the only variable that is excluded; the model also excludes these six other considerations that are not only permissible, but mandatory, for the mapmaker to consider. As a result, Mr. Trende's ensemble maps are not a representative sample of *legally compliant* nonpartisan maps. Indeed, the universe of possible maps that Mr. Trende's model created and drew its samples from includes many maps that would be unlawful. The same is likely true about some, if not many, of the 5,000 maps in Mr. Trende's ensemble. (Mr. Trende did not provide detailed information about the 5,000 simulated maps in his ensemble, so it is not possible to analyze the specific attributes of any of those maps.)

46. Mr. Trende acknowledges that his model does not consider data on race or racially polarized voting, and does not consider the need to comply with the Voting Rights Act in generating the ensemble of simulated maps. Instead, Mr. Trende examines after the fact the

number of majority-minority districts (which he calls “minority majority districts”). I am not an expert on the Voting Rights Act, but my understanding is that counting majority-minority districts is a crude and incomplete proxy for the ability of minority voters to elect their candidates of choice, and that merely counting the number of majority-minority districts does not determine whether a particular map complies with the Voting Rights Act or the New York Constitution; *see* Computational Redistricting and the Voting Rights Act, Becker et al.

47. Mr. Trende does not define how he is measuring “Minority Majority Districts,” so it is not clear if he is including any district in which there is a White voting age population of less than 50%, any district in which a single racial minority group constitutes over 50% of the voting age population, or something else.

48. Because Mr. Trende only considers the number of “Minority Majority Districts” without a more detailed analysis, my understanding is that his analysis does not accurately compare whether the ensemble maps comply with the Voting Rights Act to the same extent as the enacted maps.

49. Beyond the Voting Rights Act, Mr. Trende does not even mention the other redistricting considerations that are omitted from his model, and he makes no attempt to incorporate those considerations or to check how those omissions affect his ensemble.

Flaws in Defining the Target Distribution

50. Setting aside Mr. Trende’s failure to include a number of mandatory redistricting considerations in his model, even with respect to those he does include (equal population, compactness, and avoiding county splits), Mr. Trende does not provide enough information about his methodology to evaluate whether his model produces a representative sample of nonpartisan maps that comply with those criteria.

51. The question of whether Mr. Trende's ensemble is a representative sample of nonpartisan maps that were drawn using his specified criteria depends on how the model defines the *target distribution*. For example, if Mr. Trende were to claim that he is sampling from the *uniform target distribution*, then the question would be: does his ensemble behave as if its 5,000 maps were selected from among all possible compliant maps uniformly? That is, does the algorithm guarantee that all compliant maps are equally likely to be chosen for his ensemble?

52. Mr. Trende does not specify any target distribution at all, except to say that the ensemble is meant to replicate a scenario in which 5,000 humans were sent forth to draw maps without access to partisan information. Depending on how its parameters are set, the McCartan-Imai algorithm is capable of sampling from the uniform distribution. It is possible that Mr. Trende attempted to target the uniform distribution on "compliant" maps (with "compliant" interpreted to mean that the maps comply with the limited selection of constitutional requirements that he elected to consider: contiguity, compactness, population equality, and county preservation, without consideration of the other constitutional requirements).

53. But even this depends on the method by which Mr. Trende asked the algorithm to preserve counties. McCartan and Imai describe two choices: a slow method that requires the user to do more of the work by hand, and a fast method that functions essentially like a switch built into the algorithm's user interface. If Mr. Trende used this "county preservation switch," then this would render the algorithm incapable of also targeting the uniform distribution. When this switch method is used, then many compliant maps have zero probability of making it into the ensemble.

54. In summary, it is not clear based on the limited information that Mr. Trende has provided whether his ensemble has produced a representative sample of nonpartisan maps even with respect to the few constitutional requirements that he chose to incorporate into his model.

Flaws in Mr. Trende's Application of the McCartan-Imai Model

55. I have read the McCartan-Imai paper (on which the ensemble-generation algorithm used by Mr. Trende is based) carefully, and I think that the probability calculations are clever and elegant. But because the algorithm is so new, and because of some technical under-the-hood considerations, I believe that a lot of extra care is required to use this algorithm in redistricting litigation. Extra care is particularly required in selecting a large enough ensemble size. The McCartan-Imai paper is not yet published in a refereed journal. In fact, it very recently received a referee report that requires the authors to make some significant changes to the paper. Their paper remains a work in progress.

56. In contrast, the more commonly used *Markov Chain* algorithms were developed and tested and improved over several years by multiple teams of mathematicians and statisticians working sometimes together and sometimes in parallel. The kinks have been worked out. The foundational issues have been clearly delineated and studied.

57. One limitation of the older *Markov Chain* method is that it is typically impossible to rigorously decide how large of an ensemble one needs. Conscious of this limitation, practitioners are careful to run all kinds of validations. To verify that an ensemble of, say, 50,000 maps in a specific state is large enough for a specified purpose, practitioners can re-create the ensemble multiple times. They can create larger ensembles. They can use different starting seeds. They can verify that none of these changes affect the outcome, all of which provides strong evidence that 50,000 was enough.

58. I am very surprised that Mr. Trende constructed an ensemble of only 5,000 maps using a brand-new method, and that he apparently performed no such validations to check that 5,000 is enough. Personally, I believe that 5,000 probably is not enough for this particular application. New York has over 15,000 precincts that must be partitioned into 26 congressional districts. The larger the number of precincts and districts, the larger the ensemble that is needed. The McCartan-Imai algorithm is very different from the more established Markov Chain algorithms, and not much is known yet about the required ensemble size. There are technical reasons to expect that very large ensembles might be necessary, especially if one wishes to target the uniform distribution or any other target distribution that differs significantly from the mathematically complicated “spanning-tree” distribution, which the algorithm is natively inclined to target. Even if you knew that 5,000 maps were enough with the county-preserving switch turned off, they might not be enough with the switch turned on. This switch affects a major change to the core working of the algorithm, so all bets would be off.

59. In summary, I will not be surprised if the McCartan-Imai algorithm grows into a standard part of the redistricting analysis tool kit, but right now I think that a lot of extra care is required to use in court an algorithm that has received so little test driving. It certainly should never be used without validations to check that the ensemble size is adequate and that the ensemble is structurally sound. Moreover, it should not be used by someone who only engages its user interface at a surface level without thinking carefully about how its inner workings might affect the results in a particular case, as Mr. Trende seems to have done.

Summary of Analysis of Methodology

60. In summary, I believe with a high degree of professional certainty that Mr. Trende's methodology is flawed and that the ensemble he created is not a representative sample of lawful maps that could be drawn without partisan considerations.

RETRIEVED FROM DEMOCRACYDOCKET.COM

Dated: February 24, 2022

Kristopher R. Tapp
Kristopher R. Tapp

Sworn and Subscribed before me this 24th
day of February, 2022

Brandy M. Connor
Notary Public

My Commission Expires: *September 28, 2024*

Commonwealth of Pennsylvania - Notary Seal
BRANDY M. CONNOR, Notary Public
Philadelphia County
My Commission Expires September 28, 2024
Commission Number 1004805


RETRIEVED FROM DEMOCRACYDOCKET.COM

CERTIFICATE OF CONFORMITY PURSUANT TO N.Y. C.P.L.R. § 2309(c)

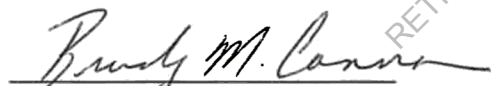
I, Michael J. Corusso, do hereby certify and attest that I am an attorney duly admitted to practice law in the Commonwealth of Pennsylvania

I make this certification for the purposes of compliance with New York State Civil Practice Law & Rules Section 2309(c) with regard to the foregoing Affidavit of Kristopher R. Tapp, to be filed in Supreme Court in Steuben County, State of New York.

Said Affidavit, acknowledged and sworn by Mr. Tapp before a Notary Public in and for the Commonwealth of Pennsylvania, and said Affidavit being therein sworn in the Commonwealth of Pennsylvania, is and appears to be, based upon my review of said document and notarization thereof, in conformity with the laws of the Commonwealth of Pennsylvania for the making of an affidavit and the notarization thereof.


PA Bar # 203684
NY Bar # 5731435

Sworn and Subscribed before me this 24th day of February, 2022


Notary Public

My Commission Expires: September 28, 2024

