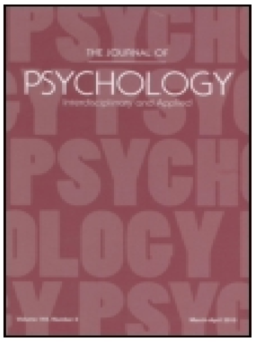

The Journal of Psychology

ISSN: 0022-3980 (Print) 1940-1019 (Online) Journal homepage: https://www.tandfonline.com/loi/vjrl20

The $P_1 P_2/D$ Hypothesis: The Case of Railway Express

George Kingsley Zipf

Published as a separate and in *The Journal of Psychology*, 1946, **22**, 3-8.

THE $\frac{P_1P_2}{D}$ HYPOTHESIS: THE CASE OF RAILWAY EXPRESS*

Harvard University

George Kingsley Zipf

In 1940 the author published the observation (5) that the following equation of the generalized harmonic series described the recent distribution of communities in India, Germany, and certain other countries including the United States (for communities of 2500 or more inhabitants), when the communities are arranged in the order of decreasing size, with A representing the population of the largest community, and with the denominators referring to the ranks of the communities thus arranged:

$$ASn = \frac{A}{1^p} + \frac{A}{2^p} + \frac{A}{3^p} + \ldots\ldots + \frac{A}{n^p}$$

For 1930 in the U.S.A. the value of p was about 1; for 1940 it was .95 approximately.[1]

In 1941 the author presented a fuller treatment (4) of the topic and included a theoretical discussion, in the form of a lemma (*ibid.* pp. 91-135), of why the proportions of the generalized harmonic series will emerge in the communities of a national social system under the postulate of reducing to a minimum the sum of all products of masses moved, when multiplied by their respective work-distances, D.

This lemma set forth the reasons for the emergence of communities at all, and also for their number, relative sizes, and locations under the above work-minimum. According to this lemma, the number, sizes, and locations of communities represent equilibria between the economies of the populations living immediately at the source of raw materials on the one hand (e.g., at the farm or at the mine pit), and in one big city, on the other hand, where all the manufacturing is done. This lemma, which we mention here only in barest outline, applies only to those social systems that produce what they consume and consume what they produce, and only under the assumption that the system is minimizing its total work in the entire movement of all materials and persons (i.e., all materials and persons will move in such

*Received in the Editorial Office on March 25, 1946, and published immediately at Provincetown, Massachusetts.

[1]To the best of my knowledge Felix Auerbach (1) was the first to point out the above rectinilearity, though he generalized incorrectly on the value, $p=1$, and ignored the parameter of area; (cf. 6, fn. 8).

amounts and directions and at such rates among the n different communities as to minimize the total work of the system).

Now if we assume, in the light of the above lemma, that C represents the total goods of production of the social system during a given interval of measurement, and that these goods are distributed approximately evenly among the members of the social system, then the relative flow of goods *into and from* any community whose population is P_1 will be in proportion to P_1, while that *into and from* P_2 will be in proportion to P_2. In short P_1 will be sending to the rest of the population, and receiving from it, the proportionate share of the flow, P_1/C, while P_2 sends-and-receives P_2/C.

If we now inquire into the exchange of materials between P_1 and P_2, we see that the former will give to, and receive from, the latter an amount that is equal to (P_1/C) (P_2/C), or $(P_1P_2)/C^2$, provided that we ignore the factor of the shortest work-distance, D, and assume that all other factors are constant.

If however we introduce the factor of the shortest work-distance, D, between communities, we can see that the flow of goods between the above two cities will be modified inversely by the length of the shortest work-distance, D, between them. Indeed, if our work-minimum be correct, then the work of transporting goods will be directly proportionate to D, with the result that the flow of goods between any two communities of population-size, P_1 and P_2, will follow the equation:

$$\frac{P_1P_2}{D} = C^2 .$$

In which D represents the shortest work-distance between them.

Now the factors, P_1, P_2, and D are empirically ascertainable for the communities of a given national social system, like the United States, and the problem of assessing the actual rate of flow between them for some classes of goods during an actual interval of measurement is by no means empirically impossible.

In the present paper we shall study the interchange, or flow, of pounds of Railway Express parcels (in less than carload lots) during the month of May, 1939, between 13 arbitrarily selected American cities[2]: (1) Boston,

[2]I express also here my gratitude to Mr. L. O. Head, President of the Railway Express Agency for providing me with these data some years ago, the presentation of which was delayed in the vain hope that normal times would soon return for further samplings. Mr. Head reported to me that the full carload express traffic (not included in the data) is devoted to unusual shipments, such as race-horses and the like, from which I infer that the present data may be viewed as a fairly complete population. He also assured me that the month in question was representative; therefore I have ignored the possibility of seasonal variation.

(2) Buffalo, (3) Chicago, (4) Cleveland, (5) Detroit, (6) Los Angeles, (7) Milwaukee, (8) New York, (9) Philadelphia, (10) Pittsburgh, (11) St. Louis, (12) San Francisco, (13) Washington, D. C.,—or 78 pairs of cities in all.

In Figure 1 we present the data graphically, with $(P_1P_2)/D$ plotted logarithmically on the abscissa[3] for each pair of cities, and with the actual amount of the interchange of express goods (less carload lots) in units of

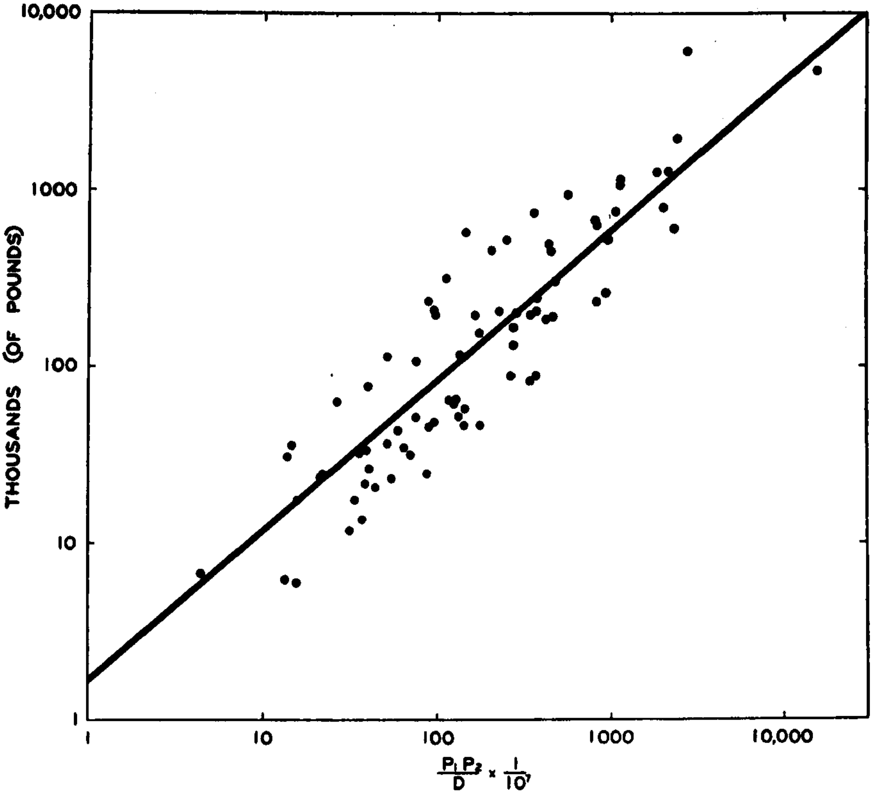

FIGURE 1

The movement of railway express (less carload lots) between thirteen arbitrarily selected cities in the U. S. A. during May, 1939, with the respective products of the populations of each pair of cities, divided by one ten-millionth of their intervening distance plotted on the abscissa.

[3]For ease of calculation and charting, the units on the abscissa refer to the values of P_1P_2 / D after being divided by the constant, 10 million, (or, multiplied by the fraction, $1/10^7$). Hence for the actual values without this simplification, the reader will imagine that there are seven empty logarithmic cycles to the left of 1, and that #1 is replaced by 10 million, and so on.

one thousand pounds plotted logarithmically on the ordinate. The values of P were taken from the official 1940 Census, and the values of D are the shortest railroad distances (the official military railroad distances) between each pair of cities.

According to our theoretical expectations, the 78 points for the 78 different pairs of cities, thus plotted, should fall on a straight line with a positive slope of 1.00 (i.e., the line will ascend from left to right at an angle of 45°).

And an inspection of the data of Figure 1 reveals that the points do approximate a straight line whose actual slope, however, as calculated by least squares, is .85 ± .31, or, in equation form: $\log y = .2157 + .8472 \log x$, (in y-units of 1,000 lbs.). According to the conventions of the trade, we may say, in view of the size of the error (the probable error being .2), that the deviation of .15 from the expected slope of 1.00 is not significant.

In any event, the linearity of the data is striking. And with variations in P from about 500,000 to 7,500,000, and in D from about 100 to 3,000, and in amounts of express from 5,000 pounds to nearly 5 million pounds, we should perhaps not be too concerned if we miss the theoretically expected slope of 1.00 by a mere .15 in plotting four variables over four logarithmic cycles.

Nevertheless, as we intimated above, our $(P_1P_2)/D$ relationship in two dimensions (like the corresponding gravitational relationship in three dimensions) may be subject to modifying factors which can significantly alter the slope and even the rectilinearity. And since many years may elapse before we again have sufficiently stable population-data to permit of a further sampling, we shall now briefly—if only for didactic reasons—inquire into the reasons for the deviation, on the *assumption* that it is significant.

If the deviation is significant, then there is a constant in the phenomenon that we have overlooked. In view of the nature of our data, the effect of the constant increases in proportion to a constant power of $(P_1P_2)/D$. That in turn can mean *among other things* either (a) that relatively *less express* is sent from larger cities than from smaller, or (b) that relatively *more express* is sent between farther cities than nearer ones.

The first alternative (a) would seem to be almost precluded by the fact that the large cities, like New York, seem to send relatively more to the more distant ones, like San Francisco, than to the nearer ones, like Philadelphia. Indeed there is nothing in the data to suggest that a person who has a parcel to send is less likely to call the Railway Express Agency if he lives in a city of 7 million instead of one of 700,000.

Hence without meaning to be dogmatic, we suggest that D, distance, is the factor, and that Railway Express becomes increasingly more attractive as a means of shipping goods as the distance of the destination increases—because, in our opinion, the Railway Express Agency, in competition with other manners of shipment, such as parcel post, is notoriously speedier. (We assume that their rates vary with the distance and are commensurate with competitive rates.)

This factor of a greater speed, or velocity, of shipment, v, will be ever more significant as D increases. Therefore, if this hypothesis be correct, the factor D should be altered to D^v, with v being a *relative* velocity in comparison with the velocities of competitive manners of shipment. This hypothesis about v, incidentally, can be tested empirically, it seems, after air express has been further expanded under reasonably free competition and more settled conditions, since, with air express, the factor of speed is of admitted importance (and we predict a slope in that case that diverges even more below 1.00).

Obviously in the above discussion we have made the implicit assumption that the ratio between goods sent by Railway Express and those sent by all other means is fixed per unit of population—an assumption that seems to be confirmed by the facts in the light of our entire hypothesis and lemma which also seems to be confirmed.

As to other empiric support of our "P_1P_2/D Hypothesis," we have observed a rectilinearity for the movement of persons by rail, by bus, by airplane, the transmission of telephone messages, telegrams, the circulation of newspapers, the places of origin of news-stories and of obituaries in newspapers, etc. Furthermore we predict that the same relationship will be found with various classes of mail, freight, the clearing of checks, etc., for which data are not available, and indeed for all movement of all sorts within a social system. These other data and relavent theoretical arguments will be presented in detail in my forthcoming book, *The Principle of Least Effort,* where the attempt will also be made to show that many isolated observations of others, such as those of S. Stouffer (3) and of J. Q. Stewart (2) are but special cases of the more primary principle, P_1P_2/D, which in turn refers to a minimizing of the total work of a bio-social system as suggested in the above-mentioned lemma.

References

1. Auerbach, F. Das Gesetz der Bevölkerunskonzentration. *Petermanns Mitteilungen*, 1913, 74f.
2. Stewart, J. Q. An inverse distance variation for certain social influences. *Science*, 1941, 89-90.
3. Stouffer, S. A. Intervening opportunities: A theory relating mobility and distance. *Amer. Sociol. Rev.*, 1940, **5**, 845-867.
4. Zipf, G. K. National Unity and Disunity. Bloomington, Ind.: Principia Press, 1941.
5. ————. The generalized harmonic series as a fundamental principle of social organization. *Pychol. Rec.*, 1940, **4**, 43.
6. ————. Unity of nature, least-action, and natural social science. *Sociometry*, 1942, **5**, 48-62.

109 Sargent Street
Newton 58, Massachusetts