Nevvtonus ab quibusdam nævibus vindicatus*

Howard Stein

Department of Philosophy
The University of Chicago

Ab quibusdam nævibus, not *ab omni nævo:* Warts and all is a good rule, and Newton did have blemishes—but not by any means all those that have been ascribed to him; and of those in some sense properly attributed, not all have been rightly diagnosed. The present paper is concerned, then, not to argue that Newton's work is without fault, but to attempt to rectify some faults of his critics.

§1. One serious and very puzzling defect appears to have been first noticed by Johannes Lohne as recently as 1961 (Lohne [1]). It concerns the experimental evidence adduced by Newton, in Experiment 15 of Book I, Part I of the *Opticks,* for the proposition that homogeneous light obeys Snell's Law—what Lohne, in the title of his paper, refers to as "Newton's 'proof' of the sine law." Lohne's main point is unquestionably correct—and it is quite amazing that the error in question should, in the first place, have been made by Newton, and, in the second, have gone unnoticed for over two hundred fifty years. On the other hand, Lohne's own analysis of the optical situation of Newton's experiment is defective, and in a way that tends to exaggerate both the magnitude of Newton's experimental error and the gravity of his theoretical lapse. There is also a point to be made—I think, one of some importance—about the nature of the claim Newton actually makes for the strength of the evidence he has obtained.

The experiment from which Newton adduces that evidence is the following: Sunlight enters a darkened chamber through a small hole, and is refracted successively by a pair of adjacent prisms. The first of these has its axis perpendicular to the midline of the incident beam; the axis of the second prism is perpendicular both to the axis of the first, and to the midline of the beam emerging from the first (which I shall call "the once-deviated beam"). The beam emerging from the second prism—the twice-deviated beam—falls upon a wall or screen, at a considerable distance from the pair of prisms, and oriented perpendicularly to the midline of that beam. Both prisms are of the same material, and each is in the position of minimum deviation: to make this stipulation

*This material is based upon work supported by the National Science Foundation under Grant No. DIR-8808575.

precise, let us assume it to hold for a ray of mean refrangibility, incident along the midline of the initial beam (the "mid-ray"). For the sake of simplicity of exposition, I shall suppose (following both Newton and Lohne) that the axis of the first prism is oriented horizontally, with its refracting edge below (so that its principal plane is vertical and the first deviation is in the upward direction); and that the midline of the once-deviated beam is horizontal (so that the axis of the second prism is vertical, its principal plane horizontal, and the second deviation of the mid-ray is in the horizontal direction—let us say, to the right). The effect of this pair of refractions by a pair of crossed prisms is to cast upon the wall a spectrum of oblong shape, in an inclined position.

In the course of the experiment, the procedure just described is repeated, always with the same first prism, but with prisms of different refracting angles in the second position. Newton ([4], pp. 78–79) reports the following (in his description, the designations P, p, $2p$, $3p$, refer to the lower (less deviated, thus red) ends; T, t, $2t$, $3t$, to the higher (more deviated, thus violet) ends of the several spectra:

> These things being thus ordered, I observed that all the solar Images or coloured Spectrums PT [the spectrum exhibited by the once-deviated beam], pt, $2p\,2t$, $3p\,3t$ did very nearly converge to the place S on which the direct light of the Sun fell and painted his round white Image when the Prisms were taken away. The Axis of the Spectrum PT, that is the Line drawn through the middle of it parallel to its rectilinear Sides, did when produced pass exactly through the middle of that white round Image S. And when the Refraction of the second Prism was equal to the Refraction of the first, the refracting Angles of them both being about 60 Degrees, the Axis of the Spectrum $3p\,3t$ made by that Refraction, did when produced pass also through the middle of the same white round Image S. But when the Refraction of the second Prism was less than that of the first, the produced Axes of the Spectrums tp or $2p\,2t$ made by that Refraction did cut the produced Axis of the Spectrum TP in the points m and n, a little beyond [that is, *below*] the Center of that white round Image S. Whence the proportion of the Line $3t$T to the Line $3p$P was a little greater than the Proportion of $2t$T [to][1] $2p$P, and this Proportion a little greater than that of tT to pP. Now when the Light of the Spectrum PT falls perpendicularly upon the Wall, those Lines $3t$T, $3p$P, and $2t$T, $2p$P, and tT, pP, are the Tangents of the Refractions, and therefore by this Experiment the Proportions of the Tangents of the Refractions are obtained, from whence the Proportions of the Sines being derived, they come out equal, so far as by viewing the Spectrums, and using some mathematical Reasoning I could estimate. For I did not make an accurate Computation.

[1] The edition cited has "or"—an obvious error for "to."

As Lohne remarks, Newton's statement that when both prisms have the same refracting angle the axis of the inclined spectrum passes through the middle of the undeviated solar image is incorrect. It is easy to see what theoretical mistake would lead to this false conclusion: one argues that the second prism deflects each ray sideways as much as the first does upwards, and that therefore (in Newton's notation) $3p$PS and $3t$TS are isosceles right triangles (with the right angles at P and T respectively); hence, since S, P, and T fall on a vertical straight line, S, $3p$, and $3t$ fall on a straight line inclined at 45°. This argument overlooks the facts (a) that the rays which (coming from the center of the solar disk) fall upon P and T do not lie in the principal plane of the second prism, so that their sideways deflection is not quite the same as their previous upwards deflection by the first prism (in whose principal plane they do of course lie); and (b) that these rays do not fall perpendicularly upon the wall at P and at T. It is indeed true of the mid-ray that its positions in the spectra PT and $3p\,3t$ form, with S, a right triangle of the indicated sort; but the same is not true for the extreme rays of the spectrum. It would appear, then, that Newton regarded the departures from the principal plane and from perpendicularity as negligible—cf. his rather loose phrase, "when the Light of the Spectrum PT falls perpendicularly upon the Wall." But since the small angular differences concerned are just what account for the elongation of the spectrum in the first place, and since the point under investigation—it being already known that the law of sines is satisfied to some reasonable degree of accuracy when dispersion is negligible[2]—is to determine whether this law is satisfied accurately for the several rays with those several small angular differences, Newton's casual treatment of this point is surprising and disturbing.[3]

To simplify geometrical considerations, Lohne discusses the behavior of those rays only that are incident on the first prism along the midline of the entering beam; and takes the two prisms to be of negligible thickness (relative to the distance of the wall on which the spectra are cast), and to cross just at the point of incidence of those rays. In effect, therefore, each incident ray is treated

[2]Newton ([4], p. 76) expresses this as the presumption that "the Rays which have a mean Degree of Refrangibility ... are refracted according to a given Proportion of their Sines."

[3]R. S. Westfall ([1], p. 213, n. 103) minimizes the importance of this mistake, calling it "a simple error arising from the geometry of the room in which Newton had to perform the experiment." This comment seems to me quite as faulty as those (having the opposite tendency) with which I am principally concerned in this paper. Surely to blame the *room* for Newton's mistake in geometry is rather hard. The whole problem is by its very nature a geometrical one, and of some delicacy; the responsibility for treating it correctly falls upon the geometer—not upon the space that forms the subject of his reasoning! Moreover, the question whether the line $3t\,3p$ passes through the point S is essentially independent of the geometry of the room. The three points in question are determined by three lines—those of the corresponding light rays—which can be treated to a reasonable approximation as diverging from a common point (for the prisms are close to one another and to the hole in the window-shutter, and far from the wall). Whether the points are collinear, therefore, depends entirely upon whether the corresponding lines are coplanar, and has nothing to do with the orientation of the wall or screen that receives the light.

as experiencing simultaneously, at that point, a pair of defelections—upwards and to the right.[4]

These simplifying assumptions are unobjectionable: the effort required for a rigorous treatment—taking account of the angular diameter of the sun, the diameter of the hole, the thicknesses of the prisms, and the distances of the latter from each other and from the hole—would hardly be repaid by the resulting gain in precision. But Lohne takes a further step that has far more serious consequences. He modifies (conceptually) the experiment itself, in a way he describes as follows (I, pp. 389–390):

> For theoretical reasons we wish equal refractions on both sides of the [first] prism, so we rotate it slightly as we determine the different parts of the spectrum.... For each separate index of refraction we must ensure equal refractions also for the second prism before we mark out where the corresponding ray impinges on the wall. Newton, who considered only a very small spectral range [namely, of course, that of visible light], did not bother about this but provided symmetry only when the index of refraction was about 1.54.

Under these conditions, Lohne obtains for the curve of the spectrum on the wall the following pair of parametric equations:

$$x = R\tan(\varphi - 37^\circ 11')$$

$$y = R\tan\varphi\sec(\varphi - 37^\circ 11').$$

Lohne does not explain the notation introduced here. It is clear, however, that x and y are intended as Cartesian coordinates on the plane of the wall; this is confirmed by Lohne's Fig. 1 (p. 389), which shows that the axes of x and y are taken to be vertical and horizontal respectively, with the origin at the point of the wall on which the once-deflected mid-ray would fall. One easily sees that $37^\circ 11'$ is what Lohne takes for the deflection of the mid-ray by the first prism; that φ is the magnitude of the deflection of an arbitrary ray; and that the deflection (of a given ray) is here assumed to be the same for both prisms (i.e., these equations are in fact derived for the case of two prisms whose refracting angles are equal). R is of course the perpendicular distance from the crossing-point of the prisms to the wall.

With these assumptions, Lohne's results are quite correct;[5] but the thought-experiment he has envisaged differs rather seriously from Newton's real one.

[4] Lohne also supposes, at first, that the chamber on whose walls the spectra are displayed is spherical, with the crossing-point of the prisms at the center of the sphere. In the analysis to be discussed here, however, this supposition plays no role—Lohne proceeds to introduce the plane wall that corresponds to Newton's actual arrangement, and his equations (Lohne [1], p. 390) and diagram (ibid., p. 389) represent the spectra on that plane.

[5] There is, however, one very odd slip in his exposition. Lohne says (ibid., p. 390)—referring first to his schematic diagram (Fig. 1—ibid., p. 389)—"The spectrum on the spherical wall will now follow the curve ADE, whose projection onto the rear wall is a *circle*" (emphasis in

First, it should be remarked that when Lohne, speaking of the "necessity" (in his version of the experiment) of moving the second prism each time a ray of different refrangibility is considered,[6] says that "Newton ... did not bother about this," his comment is a little misleading. It suggests that, although Newton, "who considered only a very small spectral range," could afford to dispense with this nicety, he might have chosen not to—might, that is, have carried out the experiment in a more scrupulous fashion. But in fact, for more than one reason, the experiment could not possibly have been executed in the manner described by Lohne. That would, in the first place, have required that the incoming beam of light be maintained in a fixed direction during the whole time of investigation—one by one—of successive rays of varying refrangibility across the entire spectrum; which, in view of the constantly changing of position of the sun in the sky, was quite impossible for Newton to arrange. Next, even if this problem could have been overcome (e.g., if Newton had invented and constructed for himself a heliostat), there would have been no way for him to *identify* the point of the doubly deflected image at which there arrived the ray corresponding to a particular inclination of the second prism. For, one must

---

the original). A glance at Lohne's equations for the curve of the spectrum is enough to show that this curve cannot be a circle, since these equations are those of an analytic parametrized curve on which both coordinates go to infinity as the parameter φ approaches $127°11'$ (and on which, furthermore, y goes to infinity while x remains finite as φ approaches $90°$).

Although it is of no importance for what follows, it seems worth noting as a curiosity that when (as Lohne's equations presuppose) the refracting angles of the prisms are equal, the spectral curve on the spherical surface initially considered by Lohne is in fact a hippopede of Eudoxus. The latter curve is, in general, the locus of a point on the equator of a sphere S' whose axis A' is fixed to the surface of a second sphere S, concentric with S', when S rotates uniformly about an axis A fixed in space, and S' rotates, simultaneously and at the same angular velocity, about the (moving) axis A'. In the optical situation envisaged by Lohne, A corresponds to the line of the first prism, A' to that of the second (which is moved as one considers, successively, rays of increasing index of refraction). It is usually specified, in discussions of the hippopede—e.g., Neugebauer [1], pp. 182–183; or Neugebauer [2], vol. II, pp. 677–678 and vol. III, pp. 1357–1358, Figs. 23–24—that the senses of rotation about the two axes are opposite. This stipulation, derived from the astronomical application made by Eudoxus, ceases to have any meaning when, as here, the axes are orthogonal. But variation of the angle between the axes generates an analytic family of analytic curves embracing both the case of orthogonality and that in which the senses of rotation are the same; there is therefore no geometric reason to limit the use of the name "hippopede" in the traditional way.

Setting aside the cases in which the two axes coincide—when, if the senses of rotation are opposite, the curve degenerates to a point; if the senses are the same, to a circle—the hippopede is a curve with a double-point (self-intersection); and this property is preserved under the central projection onto the plane of the wall that yields Lohne's spectral curve. The double-point is encountered at the parameter values $\varphi = 0°$ and $\varphi = 90°$. (Of course both these values fall far outside the visible spectrum.)

Lohne's mistake about the form of the spectral curves is repeated by Alan E. Shapiro in his edition of Newton's *Lectiones Opticæ*; and he also errs in referring to the projection of Lohne's spherical image onto the plane of the wall as a *stereographic* (rather than central) projection. (See Newton [3], p. 444, n.)

[6] Cf. the quotation already given in the text, in which Lohne says: "For each separate index of refraction we *must* ensure equal refractions also for the second prism before we mark out where the corresponding ray impinges on the wall" (emphasis added here).

remember, the initial beam is white, and what Newton saw on the wall was an entire spectrum of colors. The visual discrimination of the color would not be a sensitive enough criterion to allow the necessary identification of the ray (since rays that are near, but distinguishable, in refrangibility, will be indistinguishable in color). What one has to imagine for the execution of Lohne's experiment is, rather, this: We use, successively, incident beams of *monochromatic* light (all along the same line of incidence); we determine, for each, its deflection by a single prism adjusted to give minimum deviation; we then rotate the second prism accordingly; and finally we mark the spot on the wall to which this ray, doubly deflected, comes. Repeating this procedure for monochromatic lights through the whole spectral range, we build up in pointillist fashion the Lohne spectral curve. Quite obviously, this is a procedure far beyond realization with the means at Newton's disposal; indeed, with apparatus capable of such refined application, direct measurement of the angles of incidence and refraction for a series of monochromatic beams would have been possible, affording both a theoretically more direct and an experimentally simpler way to test the law of sines for monochromatic light. Next, even if this problem could have been overcome (e.g., if Newton had invented and constructed for himself a heliostat), there would have been no way for him to *identify* the point of the doubly deflected image at which there arrived the ray corresponding to a particular inclination of the second prism. For, one must remember, the initial beam is white, and what Newton saw on the wall was an entire spectrum of colors. The visual discrimination of the color would not be a sensitive enough criterion to allow the necessary identification of the ray (since rays that are near, but distinguishable, in refrangibility, will be indistinguishable in color). What one has to imagine for the execution of Lohne's experiment is, rather, this: We use, successively, incident beams of *monochromatic* light (all along the same line of incidence); we determine, for each, its deflection by a single prism adjusted to give minimum deviation; we then rotate the second prism accordingly; and finally we mark the spot on the wall to which this ray, doubly deflected, comes. Repeating this procedure for monochromatic lights through the whole spectral range, we build up in pointillist fashion the Lohne spectral curve. Quite obviously, this is a procedure far beyond realization with the means at Newton's disposal; indeed, with apparatus capable of such refined application, direct measurement of the angles of incidence and refraction for a series of monochromatic beams would have been possible, affording both a theoretically more direct and an experimentally simpler way to test the law of sines for monochromatic light.[7]

[7] In fact, direct measurement of the sines of incidence and refraction for monochromatic lights was itself not beyond Newton's means to achieve, and it is puzzling that he did not carry out such measurements. But the point remains: what Newton has done is to exploit an experiment contrived for another purpose to attempt to check Snell's law for the separate species of homogeneous light; to have tried to institute the experiment of Lohne for this purpose—even if it had been feasible—in *preference* to a direct measurement, would have been quite insane.

But the important question is whether Lohne's substitute experiment gives geometrical results that differ markedly from those of the actual one.[8] The detailed discussion in the Appendix below shows that there are indeed significant differences. The results of that discussion may be summarized as follows:

First, Lohne is unquestionably right when he says that Newton's diagram is erroneous, and that the spectrum formed by two identical crossed prisms does not point towards the position of the undeviated solar image; Newton has clearly made both a theoretical mistake and an inaccurate observation.

Second, Lohne and Newton are both wrong in representing the several spectra as having extremities that fall on two horizontal lines: the extremities of the spectra (in the actual experiment of Newton, in contrast with Lohne's theoretical substitute) lie on hyperbolas convex towards the horizontal line that joins the several image-points of the mid-ray.

Third, Lohne's substitute experiment appreciably exaggerates the magnitude of Newton's mistake about the direction of the line of each spectrum. For a pair of glass prisms, both having refracting angles of 60°, and with index of refraction 1.547 at the mid-ray—a value determined by Newton for one of his prisms—Lohne's results give for the slope of the spectral line the value 0.564, whereas the true value for Newton's experiment is 0.751; for water-filled prisms of that same refracting angle the corresponding slopes are 0.833 for Lohne, 0.913 for Newton (whereas Newton's claim is that the slope for two identically constituted prisms is unity).

Fourth, Lohne's diagram is also at fault (for the real experiment—not for his own substitute) in representing the spectral curves as uniformly convex towards the vertical line through the undeviated solar image; the true curves for Newton's actual experiment are convex in the optical range, but have a point of inflection before their intersection with that vertical line (i.e., for each curve, before it reaches the position of the undeviated image). A consequence is that, for Lohne's curves, the line of the visible spectrum (that is, the tangent line to the curve at a point in the visible range—say at the mid-ray) always passes *above* the position of the undeviated image. On the other hand, for the true curves, in the case of water-filled prisms, although the line of the spectrum formed by identical prisms passes above the undeviated image (as the result given above for its slope shows), for smaller refracting angles of the second prism this line *can* pass *below* the undeviated image, as Newton represents it as doing; for example, it does so when the refracting angles are 60° for the first and 30° for the second prism.

If we ask how far these results go towards "vindicating" Newton, the answer is: at most moderately far. The theoretical mistake is clearly serious and irreparable. As to the mistakes in observation, that about the lines through the extremities of the spectra is surely a venial one—for the extremities of the

[8] Shapiro (in Newton [3], pp. 443-444, n. 4) follows Lohne's analysis of Newton's error, with no notice of the change that has been made in the conditions of the experiment. Presumably both Lohne and Shapiro have assumed that the change is of negligible effect.

spectra are (as Newton remarks) not sharply defined, "the Light there decaying and vanishing by degrees" (Newton [4], p. 29). The mistake about the slope—or the claim that the line of the spectrum produced by identical prisms passes through the position of the undeviated image—remains considerable for the case of glass; but Alan Shapiro has pointed out that Newton himself recommends the use of hollow water-filled prisms for this experiment, since he found it easier with prisms of this type to control the refracting angles with some precision.[9] The true slope in this case—0.913—is tolerably close to 1; and if we consider that the actual spectrum is not a line, but a figure of appreciable width, the possibility of an error of that magnitude in determining the slope is increased. Thus Newton's mistake in observation becomes, if not (as in the former case) quite *venial,* at least more nearly *understandable.*

But none of these considerations meets the really crucial issue, namely: *How could Newton claim to have tested so important—and delicate—a proposition, by a theoretical analysis as cursory as must have been that which left him in such an error, and by observations as imprecise as, at best, the foregoing analysis shows his to have been?* Lohne's own conclusion is formulated ([1], p. 391) in the words: "Newton's diagram is erroneous.... The proportions [of the sines] cannot come out equal from Newton's diagram." We know that Lohne is right in the first of these assertions, even if he overestimates the magnitude of the error. What is to be said about the second assertion, and about the whole quality of Newton's argument?

I have said earlier that there is a point of some importance to be made about the nature of the claim Newton makes for the strength of the evidence he has obtained. We shall see that there is also a point of some importance about the evidence itself. For Newton does not in fact say that the proportions of the sines come out equal *from his diagram*; I have already quoted his own statement (Newton [4], p. 79):

> [B]y this Experiment the Proportions of the Tangents of the Refractions are obtained, from whence the Proportions of the Sines being derived, they come out equal, so far as by viewing the Spectrums, and using some mathematical Reasoning I could estimate. For I did not make an accurate Computation.

If it were not clear enough from this that Newton himself regarded his result, although plausible and fairly convincing, as less than decisive, it would certainly become clear from the fact that he immediately proceeds to offer a second, purely theoretical argument, that is explicitly (and quite contrary to his usual procedure—and methodological creed) *conjectural.* He says (ibid.; emphasis in original):

[9] Newton's words ([4], p. 78) are: "But for want of solid Glass Prisms with Angles of convenient Bignesses, there may be Vessels made of polished Plates of Glass cemented together in the form of Prisms and filled with Water."

> So then the Proposition holds true in every Ray apart, so far as appears by Experiment. And that it is accurately true, may be demonstrated upon this Supposition. *That Bodies refract Light by acting upon its Rays in Lines perpendicular to their Surfaces.*

(Let me add that, in my own judgment, although I have said that Newton here makes an exception to his own methodological rule, he is not guilty in this of what he calls "feigning hypotheses." For there is no feigning involved: he is quite frank about the *conditional* character of his argument, whose tenor is that a conclusion for which he has been unable to give entirely satisfactory experimental evidence will hold strictly, *if* a certain not implausible supposition is true. He does not claim that this theoretical—hypothetical—supplement to the experimental evidence clinches the case for that conclusion.)

But now to the question of Newton's actual argument from the experiment, which I have said is not based on his diagram of the skew spectra. He says that from his experiment he determined "the Proportions of the Tangents of the Refractions," and from these "the Proportions of the Sines"; and that the latter "come out equal, so far as by viewing the Spectrums, and using some mathematical Reasoning I could estimate." Just what does this mean?

Two facts are immediately evident. In the first place, from the ratio of the tangents of two angles it is not possible to determine the ratio of their sines. In the second place, the angles the ratio of whose tangents Newton has measured are the *total angular deflections of the several rays by the second prism*; but the angles the proportions of whose sines it is pertinent to estimate are *the angles of refraction of the several rays at the same angle of incidence.*[10] Thus Newton's phrase "the Tangents of the Refractions" is loose, and his characterization of the whole procedure of estimation is still more so. Unless and until the study of his unpublished papers turns up material relevant to the question of what his actual measurements and his actual "mathematical Reasoning" here may have been, we can only proceed by reasonable conjecture.

I have been able to form only one reasonable conjecture about this. Newton has told us (not, to be sure, what his data were, but) what he measured: he says "those Lines $3t$T, $3p$P, and $2t$T, $2p$P, and tT, pP, are the Tangents of the Refractions." These are horizontal lines connecting points of the several inclined spectra with the central, vertical (once-deflected) spectrum; and Newton can only have taken them, as I have said, for the tangents of the angles of total deviation by the second prism. They are in fact, on the more accurate analysis given in the Appendix below, (proportional to) the tangents of the angles of total deviation of what are there called the *projected rays*. The error committed

[10] That is, since Newton assumes the correctness of Snell's law for some particular ray (of "a mean Degree of Refrangibility"—cf. n. 2 above), its correctness for all other rays will have been established if he can show that the ratios of the sines of refraction of rays of given spectral species at the same angle of incidence remain the same as one varies that angle of incidence.

in substituting the projected deviations for the true ones is easily seen to be entirely negligible.

There is a second error; for the corresponding (projected) rays whose tangents are so measured for the several spectra (i.e., for the several refracting angles of the second prism) do not represent quite the same species of homogeneous light. Indeed, the rays that are taken by Newton to correspond are those that fall on the same horizontal line, whereas the analysis given in the Appendix below shows that the true locus of termination of rays of the same species is a hyperbola. Again, however, the errors here are very minute: the small discrepancy that produces a detectable but slight difference in the *vertical* positions of the several rays—*all* the vertical displacements being fairly small—gives rise to a quite inappreciable difference in their horizontal displacements from the central (i.e., once-deviated) spectrum; or, more precisely, a difference that has an inappreciable effect upon the ratios of those horizontal displacements. In short, although Newton has committed errors of theory in his *precise* interpretation of what he has measured—and once more it should be remembered that he has acknowledged, in general terms, that he did not make "an accurate Computation"—these errors are in fact of insignificant quantitative effect.

It remains to consider how from these "Tangents of the Refractions" the "Proportions of the Sines" can have been derived. I have been able to think of only one way. The position of each prism, we know, was that of minimum deviation for the mid-ray; and therefore near the position of minimum deviation for every visible ray. The total deviation has a stationary value at the position of minimum deviation; it is therefore hardly affected by a slight departure from that position. But in the position of minimum deviation, the two angles within the prism—the angle of refraction of the entering ray and the angle of incidence of the emerging ray—are equal to one another: each is equal to half the refracting angle of the prism; and the total deviation is twice the angle of incidence, minus the refracting angle. From the tangents of the total deviations, therefore, the angles of incidence at minimum deviation can be obtained.[11] And these are angles of incidence (for the several rays) that correspond to equal angles of refraction within the prism—or, looking at the emerging rays instead, they are

[11] Note that I here assume Newton to reason, not from "the Proportions of the Tangents of the Refractions" (as he puts it), but from the tangents themselves. In defense of this interpretation there are two points to be made. The principal one has already been suggested: the ratios of the tangents of these angles simply do not suffice to determine the quantities of interest. But then how is one to understand Newton's own phrase? The second point bears upon that question. It was most usual, in Newton's time, for the terms "tangent," "sine," etc., to be referred primarily to *arcs* rather than angles, and to be used for certain *lines* and their *lengths,* rather than (as now) for associated *ratios* (and thus pure numbers). And indeed, Newton does here identify *the horizontal line-segments from the central spectra to the oblique ones* as "the Tangents of the Refractions." In this usage the angle is functionally associated, not with the tangent, but with "the ratio of the tangent to the radius." If Newton had said that these ratios (or "the proportions of the tangents to the radius") are obtained from the experiment, all would be clear. I am in effect taking him to use the phrase "the Proportions of the Tangents," casually and loosely, to mean essentially this.

angles of refraction (into the air) that correspond to equal angles of incidence (within the glass) for the several rays. The sines of these angles can thus indeed be determined from measurements of exactly the kind that Newton intimates. If Newton followed the procedure I have here suggested, he can perfectly well have obtained results which (a) looked compatible with a constant ratio of the sine of incidence to that of refraction "of every Ray considered apart," and (b) *were,* if interpreted on the basis of an accurate theoretical analysis, quite genuine evidence in favor of the conclusion that that ratio is really constant.

I think this not only goes far to vindicate Newton's argument, but also suggests a quite plausible explanation for the otherwise amazing series of mistakes that Newton did in fact commit. For if his evidence for a rather basic proposition depended upon those aspects of his exposition that are erroneous, it is really baffling to understand both how he could here have been so careless (whereas, in Lohne's words—[1], p. 391—"[h]e was ordinarily very painstaking and accurate"), and how—having made those mistakes—he could have thought he had results confirming the proposition (cf. again Lohne's statement, "The proportions cannot come out equal from Newton's diagram"). But the procedure I have suggested would, as I have said, yield genuinely confirming results; and furthermore, since this procedure is quite independent of the negligent parts of Newton's analysis and observations, it is not nearly so hard to see how he could have been so lax with respect to those.

It is of course possible that Newton's procedure was different from the one I have here proposed; and it is much to be desired that evidence on this point may be found among his manuscripts that will clarify what he meant by the phrase "the Proportions of the Sines ... come out equal, so far as *by viewing the Spectrums, and using some mathematical Reasoning* I could estimate." But it is at any rate clear that he must have done *something* that is referred to but not made explicit by that phrase, and not indicated by his diagram. If, then, the course I have suggested is one he *could* have followed, and one which, notwithstanding his errors of theory and observation, would have provided genuine evidence for his conclusion, it remains possible that whatever alternative procedure he used shared this character.

Yet a little more deserves to be said here. Lohne, having made his quite devastating evaluation of "Newton's 'Proof' of the Sine Law"—summed up in the words, "Seldom has a physical law been 'demonstrated' by experiments so inaccurate and by deductions so faulty"—proceeds to extenuate the fault with the remark ([1], p. 391): "But what of it? There is such inner consistency in the circumstantial evidence presented elsewhere in the *Opticks* by Newton that most of us are convinced even without direct proof." This, however, is specious; and it is a merit of Newton's to have recognized that. What most of the evidence presented in the *Opticks* (and in Newton's other writings on the subject) goes to show is that the rays of light have diverse "refrangibilities," and that rays of any given refrangibility are refracted according to *some* definite law (close—as he says, for rays of "a mean Degree of Refrangibility"—to Snell's law). This does not at all establish—or even begin to confirm, by any appeal to "inner consistency in the circumstantial evidence"—that the accurate law for homogeneous light *is* that of the constant ratio of sines. Newton puts it with his usual clarity ([4], p. 75): "That every Ray consider'd apart, is constant to itself in some degree of Refrangibility, is sufficiently manifest out of what has been said"; moreover (ibid., p. 76), given the evidence that Snell's law holds for the "mean" rays, the satisfaction of just that law by each species of ray considered separately "is very reasonable"; "but," he adds, "an experimental Proof is desired." He does *not* claim that the evidence he has assembled in its entirety lends "circumstantial" support to this conclusion; and it in fact does not.

However, there is one piece of experimental evidence in the *Opticks* that does lend further support to the proposition, and Newton points this out. In Book I, Part I, Proposition VII Newton discusses at considerable length the chromatic aberration of lenses, and presents a good deal of relevant data. In the course of that discussion, he makes use of data on the dispersion of the extreme rays of the spectrum taken from his prism experiments, and compares the results of those experiments with the differences he determines for the focal length of a lens in homogeneous lights of those extreme varieties. The angles of incidence in the prism experiments are (of course) quite large; the angles of incidence upon the lenses are (of course) very near zero. Newton's observed values for the focal lengths agree well with the estimates he makes from the spectra—estimates that depend upon Snell's law for the several sorts of rays, each by itself. After presenting his results, he adds the remark ([4], p. 93): "And this is a farther Evidence, that the Sines of incidence and Refraction of the several sorts of Rays, hold the same Proportion to one another in the smallest Refractions which they do in the greatest." I think it fair on the whole to conclude that, notwithstanding his mistakes, Newton proves to have been both penetrating and responsible in his treatment of this question.

§2. Two other matters discussed by Lohne in the same paper, although far less important, call for some comment.

The first of these concerns the expected shape of the solar image, in Newton's initial observation with a single prism, on the basis of the received law of refraction. In his letter of February 6 1671/2 to Oldenburg containing his "New Theory about Light and Colors," Newton expresses his "surprise" at seeing the "vivid and intense colours" produced by refraction through the prism "in an *oblong* form; which, according to the received laws of Refraction, I expected should have been *circular.*" (Newton, [1], p. 92; [2], p. 48). He gives in this place no account of the reasoning that leads to this expectation. However, he does inform us that he had arranged his prism so that "the Refractions on both sides the Prisme, that is, of the Incident, and Emergent Rays, were as near, as I could make them, equal," and that the rays fell perpendicularly upon the wall ([1], p. 93; [2], p. 49); and this is enough—given that the symmetrical arrangement of the rays makes for a stationary value of the deviation, and that the angular differences among the incident rays (forming as they do a solid right circular cone whose generators are inclined $1/4°$ to its axis) are small—to allow one to conclude that the image will differ little from a circle. Of course, by *how* little they will differ is a question that calls for a more careful estimate. But in place of a theoretical argument—and quite independently of the "received laws of Refraction" themselves—Newton offers an *experimental* argument ([1], pp. 93–94; [2], pp. 49–50):[12]

> But because this computation [namely, one he has just reported, without giving particulars, which led him to conclude that the emergent beam should have had the same angular divergence as the incident beam] was founded on the Hypothesis of the proportionality of the *sines* of Incidence, and Refraction, which though by my own & others Experience I could not imagine to be so erroneous, as to make that Angle but $31'$, which in reality was 2 deg. $49'$; yet my curiosity caused me again to take my Prisme. And having placed it at my window, as before, I observed, that by turning it a little about its *axis* to and fro, so as to vary its obliquity to the light, more then by an angle of 4 or 5 degrees, the Colours were not thereby sensibly translated from their place on the wall, and consequently by that variation of Incidence, the quantity of Refraction was not sensibly varied. By this Experiment therefore, as well as by the former computation, it was evident, that the difference of the Incidence of Rays, flowing from divers parts of the Sun, could not make them after decussation diverge at a sensibly greater angle, than that at which they before converged.

This most elegant argument leaves hardly anything to be desired. At most, one could object that to establish the preservation of angles in *all* directions, the experiment of jiggling the prism should have been carried out by rotating it, not only about its own axis, but also about axes of other orientations.

In Newton's Cambridge lectures, however, he gives a more detailed geometrical discussion of this matter ([3], pp. 52/53–60/61).[13] He considers an arrangement in which the prism is placed between the sun and a small hole (treated as punctiform) in the window-shutter, with the refracting edge of the prism horizontal and below (so that the deviation of the light is upwards). Assuming Snell's law to hold strictly, he chooses a position of the prism that makes the angular divergence of the rays in the principal plane *exactly* equal to that of the

[12] I follow the text in [1], that of a transcript of Newton's letter made for him by his copyist Wickins; the text in [2], which is that of the letter as published by Oldenburg, differs slightly from this (namely, by omission of the phrase "& others").

[13] Newton [3] gives the Latin text on even-numbered pages, with English translation on the facing odd-numbered pages; my page references in the form "$2n/2n+1$" are to be interpreted accordingly.

incident rays (it is easy to see that this can be done). He then proves (in effect) the theorem represented by Equation (c) in the Appendix below; from which it easily follows that the angle between two rays that fall in a plane parallel to the refracting edge and make equal angles (on opposite sides) with the principal plane will be the same after emergence from the prism as before incidence upon it.[14] If, then, one considers generators of the incident cone of rays, first in the vertical plane through the axis of the cone, and then in the plane through that axis parallel to the refracting edge, one sees that the angle of the original circular cone is preserved in both cases after traversal of the prism. Newton now argues that if the ray to the center of the solar image on the wall strikes the wall perpendicularly, the horizontal diameter—i.e., the largest horizontal section—of that image will actually fall very slightly below the center. His reason is that the rays from what we may call the "horizontal extremities" of the solar disk will be "refracted a little less" than those from its vertical extremities (where "refracted a little less" must be taken to refer to the total *vertical* deviation of those rays—i.e., to what in the Appendix I have called the deviation of the projected rays). It will then follow that the rays in question, since the plane in which they lie is not quite perpendicular to the wall, will travel very slightly further than the vertically extreme rays; hence that, since they have the same angular divergence as the latter, their distance apart at the wall will be very slightly greater than that of the vertical extremes. The result, then, of this delicate (but, as we shall see, not quite correct) geometrical argument is that the horizontal diameter of the expected image falls slightly below the center of the image and is slightly longer than the vertical diameter—so that, in particular, the image will be a little wider than it is high.

Lohne comments on all this as follows ([1], pp. 396–397):

> In later life Newton was inveigled in a controversy about the shape of the earth, whether the globe is *oblong* or *flattened.* In Newton's optical polemics a similar question was raised about the sun's image, when cast on a wall by a prism. In *Lectures 1* Newton demonstrated that with *uniform rays* this solar image should be circular or slightly *flattened.* Newton's handling of this proof is characteristic as showing us amongst other things how he tried to evade difficulties. We shall therefore examine it in some detail:
>
> *1.* He imagines a symmetrical position of the prism. In this position the sun rays suffer the same refraction at the entrance as when they leave the prism.
>
> *2.* As only one single ray can be exactly symmetrical, it seems convenient to assign symmetry to the ray from the sun's center. But

[14] More exactly, this follows from Equations (a), (b), (c), together. Newton does not establish (a) and (b). However, appeal to the latter can be bypassed by remarking that the symmetry of the situation of the two incident rays with respect to the prism is enough to guarantee that the corresponding projected rays are equally refracted; and Newton clearly relies, tacitly, on this consideration.

> Newton says that the rays from the sun's upper rim shall deviate from the ideal symmetry exactly by the same angle as the rays from the lower rim. The ray from the sun's center cannot be exactly symmetrical in this arrangement.
>
> *3.* In his proof Newton supposes the prism to be on the *outside* of the hole through which the beam enters. This is peculiar, for in the experiment Newton puts the prism on the *inside.*
>
> *4.* Newton applies only pure geometry in his proof. His predecessor Harriott had calculated the deviations numerically.
>
> I think such peculiarities very informative about Newton, if they can be explained:
>
> *1.* A symmetrical position of the prism simplifies enormously both the experiment and the theoretical deductions.
>
> *2.* and *3.* I have investigated for myself the arrangement I consider natural, namely prism on the inside and exact symmetry for the ray from the sun's center. I found a slightly *oblong* or rather egg-shaped image of the sun. But if the prism is supposed to be on the outside, the solar image will be flattened, as Newton contends.
>
> The issue is somewhat confused, because it was not necessary for Newton to resort to a trick here. When we look up the corresponding place in the *Opticks* (Fig. 13), we find the prism on the inside.
>
> *4.* When we use only pure geometry it is very difficult to visualize in detail how and how much the skew rays deviate from symmetry. Newton's proof holds, as far as it goes, but it can scarcely have convinced the average student in Cambridge. For one of his assertions Newton gives no proof, "utpote nimis longam & proposito meo non omnino necessariam".

Lohne's remark that "the issue is somewhat confused" concerns his second and third heads; but, indeed, the whole issue is more than a little confused.

Perhaps the main point that needs to be emphasized is that Newton's Latin phrase quoted by Lohne—"as being excessively long and not altogether necessary to my purpose"[15]—can really be applied to the whole discussion: Newton has in fact given a more elaborate analysis than is appropriate to the experimental conditions. The difference from circularity in the shape of the image is in any case very slight—too slight to be observed in the experiment in question, where because the hole is (of course) of measurable size (if it were too small, one would have diffraction to contend with!), and because no lens is used to

[15] The phrase is slightly different in both its occurrences in Newton [3]—that is, in the early draft of Newton's lectures, given by Shapiro under the title *Lectiones Opticæ,* and the later revised manuscript (the basis for the edition of 1729, published under that same title), which Shapiro calls *Optica* (the manuscript itself bearing the title *Opticæ,* Pars $1^{a}/2^{da}$; see Newton [2], p. xv). In each of these places, the words are *utpote longiusculam* rather than *utpote nimis longam* (ibid., pp. 58, 290). The difference presumably is an editorial revision made in the edition of 1729 (Shapiro records only the more significant of such variants).

focus the image, the boundary of the image is not sharp. That is all that really matters; the more precise estimate of the shape is at most an interesting exercise in geometrical optics.[16] Nevertheless, since Lohne has made these comments, they deserve to be examined.

Lohne's introductory remark about the controversy over the oblateness or prolateness of the earth is an unhappy one. The issues involved are very different, and that concerning the shape of the earth—unlike the one under consideration here—was of quite fundamental importance.[17] Moreover, contrary to Lohne's intimation, the demonstration in *Lectures I* was in no way concerned in "Newton's optical polemics"; those lectures were not published until after Newton's death, and there is in the discussions of the 1670's no trace of Newton's claim that Snell's law implies a foreshortened image.

Next, it is surely misleading merely to say, as Lohne does in his comment on his point *1,* that a symmetrical position of the prism "simplifies the experiment and the theoretical deductions"—and still more so to imply that this is an instance of how Newton "tried to evade difficulties." For the issue posed by Lohne *presupposes* that we are dealing with a situation in which the image to be expected from Snell's law is nearly circular; and for that, a(n approximately) symmetrical position of the prism is *necessary.*

As to point *2,* Newton's choice of exact specification of the position of the prism is obviously designed to make the angular divergence of the beam in the principal plane of the prism exactly the same after as before refraction. Why Lohne says that "it seems convenient" to him "to assign symmetry to the ray from the sun's center" is (to me) most unclear. But Newton's own comment on the exact specification of the position of the prism is the right one: he says ([3], pp. 58/59–60/61), "Moreover, even if the prism's position were other than I have described, as long as the rays do not undergo a particularly unequal refraction on each side, the shape of the image will nonetheless hardly change because of that."[18] And this comment is explicitly extended by Newton to the

[16] One other remark is pertinent here. Setting aside all theoretical subtleties, it is an entirely straightforward matter, in Newton's experiment, having measured the width of the spectral image and its distance from the hole in the window-shutter, to verify that the width agrees with what is to be expected if the lateral divergence of the beam of light has been preserved at approximately $1/2°$. The measurements given by Newton in his letter to Oldenburg (Newton [2], p. 93; [3], p. 49) show that this is indeed the case. Therefore the only point that could be in reasonable contention is what *length* was to be expected for the image—i.e., what the *vertical* spread should have been. But Newton's treatment of *this* in his geometrical discussion is perfectly straightforward and elementary; so—again—it is hard to see why Lohne should think there is any serious difficulty here at all.

[17] For, in the first place, the oblate shape of the earth is a consequence of Newton's dynamics and the hypothesis of the earth's rotation—thus it is a crucial point both for his physics and for his cosmology; and in the second place, it is on the basis of that shape that Newton's dynamics and theory of gravitation together lead to a correct prediction concerning the precession of the equinoxes: for a prolate earth, the theory predicts a motion of the equinoctial points in the reverse direction to the actual one; so, in this prediction again, both the physics and the cosmology are at stake.

[18] A little care is needed in interpreting Newton's idiom here: the point is not that the image

subject of Lohne's point *3*: he continues, "Nor does it make much difference whether the opaque body EG, perforated with the hole F for transmitting the rays, is placed on the near or the far side of the prism; nor does the shape of the hole [which Newton has supposed circular] matter much, provided that it is small. For such small alterations will scarcely change the image more than a tenth or perhaps[19] a fifth of its diameter, as will be clear to one who considers it. Finally, to put all in a few words, it is plain that generally the sun's refracted image must be sensibly nearly circular, provided that in the same medium the refraction at the same incidence is always the same."

It is worth noting, with respect to Lohne's remarks "it was not necessary for Newton to resort to a trick here" and "[w]hen we look up the corresponding place in the *Opticks* ... we find the prism on the inside," that Newton's diagram of the actual experiment in the *Lectiones Opticæ,* too, shows the prism on the inside (Newton [3], pp. 50, 284).[20] The "trick," then, was for geometrical purposes solely; it was in no way a deceit. And the passage just cited from Newton explicitly remarks that this trick makes no important difference in the result.

That it was "not necessary" for Newton to resort to a trick is a questionable judgment (unless one takes *necessary* in a pedantically precise sense). There is in fact a very good reason for basing the mathematical analysis of this matter on a position of the prism outside the hole. For let us consider the combined effect upon a bundle of rays—treated here, of course, as having all the same degree of refrangibility—of refraction through a prism and passage through a (punctiform) hole. The total deviation of any ray in the bundle by its passage through the prism will be determined by its direction of incidence; thus, the prism effects a mapping from the set of incident directions to the set of emergent directions. If that transformation of directions occurs before passage through the hole, the effect of the hole is to select out a *stigmatic* beam—a bundle of rays *diverging from a common point*—having just the directions derived from the set of incident directions by the prismatic transformation. (Note in particular that this resultant beam will be strictly independent of the location of the prism—always assumed to be oriented according to Newton's stipulation—so long as that location is between the sun and the hole.) On the other hand, if the incident beam passes through the hole first, its subsequent passage through the

---

will be nearly circular unless the refractions are *very* unequal, but that it will be so as long as the refractions are *nearly equal.* (Idiomatically, of course, to say that something is "not very large" is not merely to deny that it is very large, but to assert that it is small.)

[19] Shapiro translates "or even"; the Latin is *forte vel,* and "perhaps" seems clearly the better rendering (since a fifth is less favorable than a tenth).

[20] I have not checked Horsley's edition (1782) of Newton's works—the edition cited by Lohne—to verify that the diagram in question occurs there. However, the diagram in Newton [3], p. 284, is that of the manuscript from which the published text of 1729 derives; the latter is the source for Horsley; and since this diagram is referred to quite circumstantially in Newton's text, it may reasonably be presumed that it appears in both those published editions. I do not understand why Lohne should leave the impression that only in the *Opticks* did Newton depict the arrangement he had actually used.

prism will destroy its stigmatic character: we shall have the same collection of emergent *directions,* but there will be no em common point from which all the emergent rays diverge (no punctiform virtual image). Furthermore, in this case, *the precise geometric constitution of the emergent beam will depend upon the location of the prism*; for one easily sees that the astigmatism will be the more pronounced, the further the prism is from the hole. More precisely, the character of the emergent beam will depend both upon the distance of the prism from the hole and the thickness of the prism itself at the place where the rays traverse it. (If the prism is very near the hole, we may suppose it to be very small; but at increasing distances from the hole, where the beam has diverged to more or less appreciable breadth, the prism has always to be taken to be at least large enough to accommodate the beam.)

These considerations not only serve to motivate Newton's "trick," they also raise a perplexing question about Lohne's claim concerning the effect of the actual arrangement (with the prism inside the room). He does not tell us, when he says "I have investigated for myself the arrangement I consider natural," whether this was a theoretical or an experimental investigation. If the latter, it would be crucial to know the degree of homogeneity of the light that was used; for a "slightly oblong" image, such as Lohne says he "found," might be produced by "slightly heterogeneous" light (and there is no such thing in reality as entirely homogeneous light). In either case, it would also be important to know something of the details of the geometrical arrangement. For the following is clear: (1) The variation of the character of the image with the relative positions of hole, prism, and screen, is *continuous.* (2) If, in a *theoretical* analysis, we take the prism to *coincide* in position with the hole, then the result will be a *stigmatic* beam, exactly as when the prism is outside. (3) Therefore, if the prism is inside and close enough to the hole, the image—*flattened,* as we know, for the prism outside—will still be flattened: Lohnes implication that the image is oblong when the prism is inside and the ray from the suns center refracted symmetrically cannot be right without qualification.—In view of the preceding discussion, this point is of course quite without importance; *except* as showing that the standard of accuracy that has been applied in all this to Newton is one that his critic has by far not imposed upon himself.

§3. The remaining point to be considered in this paper of Lohne concerns the "quadratic dispersion law" of Newton [3] (pp. 198/199–200/201, 334/335–338/339).

Appendix: The path of a ray of light through Newton's crossed prisms.

We must consider a ray incident upon a prism along a line not necessarily in a principal plane of that prism. To treat the deflection of such a ray, it is useful to consider, besides the true path, the orthogonal projection of this path in the principal plane. Let us refer to the segments of the projected path as the *projected rays,* and to the corresponding angles as the angles of incidence and refraction of the projected rays at the first and second refracting faces of the prism. By "refracting angle" will be meant, as usual, the angle between the refracting faces. It will be assumed in all that follows that the whole configuration is such that, in the principal plane, the refracting angle and the projected entering ray lie on opposite sides of the normal to the first face; and analogously for the projected emerging ray. This is of course the usual configuration, and it obtains for all the rays in Newton's experiment. With this assumption, no care is necessary about the signs of the angles of incidence and refraction of the projected rays—they may be taken all to be positive acute angles.

Besides the angles of incidence and refraction of the projected rays, we shall have to consider the angles between the incident and refracted rays and their respective projections—that is, the angles between those (unprojected) rays and the principal plane. Here it will be necessary to distinguish between positive and negative angles. For this purpose, we suppose that one of the half-spaces bounded by the principal plane is distinguished as the "positive" one; and we say that the angle between any ray and its projection is positive if the direction of the ray itself is from the negative towards the positive half-space bounded by the principal plane.—In all this, the term "principal plane" may be understood to refer to any member of a whole family of parallel planes; then the "positive" half-spaces corresponding to the planes of this family are to be chosen *coherently* (in a sense that should be obvious).

To obtain the relations governing the angles of interest, sufficient to determine the emergent direction of a ray of any given incident direction (under the restriction already noted to the "usual" configuration), it will be convenient to make use of the formal conceptions of Newton's particle theory of light: let us represent a ray traversing a homogeneous medium by its "Newtonian velocity-vector"—that is, by a vector whose direction is that of the ray itself, and whose magnitude is proportional to the index of refraction of the medium (and thus to the *reciprocal* of the wave-theoretic phase velocity); we may call this simply the *ray-vector.*

Suppose, then, that the incident ray-vector is $\vec{u}$. The refracted ray-vector $\vec{v}$ is determined[21] by the conditions:

(i) $\vec{v} - \vec{u}$ is orthogonal to the refracting surface;

(ii) $\langle \vec{v} | \vec{v} \rangle = \langle \vec{u} | \vec{u} \rangle + W$, where W is constant for rays of a given refractive index;

(iii) $\vec{u}$ and $\vec{v}$ are both oriented in the direction from the medium of incidence to that of refraction.

(In the mechanical interpretation in terms of the corpuscular theory of light, W is proportional to the work done on the light corpuscle as it traverses a

[21] As Newton has shown: see *Opticks,* Book I, Part I, Prop. VI ([4], pp. 79-82); cf. also *Principia,* Book I, Sect. XIV, Prop. XCIV.

conservative force-field, oriented orthogonally to the refracting surface and concentrated in an "infinitesimal neighborhood" of that surface.) Let us resolve the ray-vector $\vec{u}$ into components:

$$\vec{u} = \vec{x} + \vec{y} + \vec{z},$$

with $\vec{x}$ normal to the refracting surface, $\vec{z}$ normal to the principal plane, and $\vec{y}$ orthogonal to both $\vec{x}$ and $\vec{z}$. The corresponding resolution of $\vec{v}$ will then have the form:

$$\vec{v} = \vec{w} + \vec{y} + \vec{z};$$

here $\vec{w}$ has the same sense as $\vec{x}$, and its magnitude is determined by:

$$\langle \vec{w} | \vec{w} \rangle = \langle \vec{x} | \vec{x} \rangle + W.$$

Now, the projected incident and refracted rays are $\vec{x} + \vec{y}$ and $\vec{w} + \vec{y}$, respectively. But these, in view of the conditions that determine $\vec{w}$, likewise satisfy the Newtonian conditions for an incident and refracted ray-vector—with, of course, a modified index of refraction. The true index of refraction n is the ratio of the magnitude of the refracted ray-vector to that of the incident ray-vector; if we express the magnitude of the vector $\vec{u}$ by the corresponding letter u, and analogously for $\vec{v}$, and if we represent the magnitudes of the corresponding projected ray-vectors by u' and v', then we shall have:

$$\frac{n'}{n} = \frac{v'/u'}{v/u} = \frac{v'/v}{u'/u}.$$

But u'/u and v'/v are just the cosines of the angles between the rays (incident and refracted respectively) and their projections.

Denoting by i and r the angles of incidence and refraction of the projected rays, and by p and q the angles between the true rays and their projections (incident and refracted, respectively), we thus have:

$$\sin i = n' \sin r; \tag{a}$$

$$n' \cos p = n \cos q. \tag{b}$$

The angles i and p are given. Conditions (a) and (b) will allow us to determine r if we also know q. But from our component resolutions of $\vec{u}$ and $\vec{v}$ and their respective projections, it follows that $\sin p = z/u$, $\sin q = z/v$; and from this we see immediately that:

$$\sin p = n \sin q. \tag{c}$$

The conditions at the second refracting face of the prism correspond in the obvious way to those at the first. In particular, the angle between the ray within the prism and the principal plane remains the same—q; and from the analogue of condition (c) it is clear that the angle between the emergent ray and that plane will once again be p.

Now we have to consider the passage of a ray through the pair of crossed prisms. Since we are taking account only of rays propagated along the midline of the entering beam, and since that line falls in the principal plane of the first prism, the distinction between true and projected ray is required only for the second prism. Let the angles of incidence and refraction into and out of the first prism be α, β, γ, δ, respectively; let the index of refraction (of the material of the prism relative to air) for the ray under consideration be n; let the refracting angle of the prism be A; and let the total deflection of the ray by the first prism (that is, the angle between the entering ray-vector and the emerging ray-vector) be Δ. Then we have:

$$\sin\alpha = n\sin\beta; \tag{1}$$

$$\beta + \gamma = A; \tag{2}$$

$$\sin\delta = n\sin\gamma; \tag{3}$$

$$\Delta = \alpha + \delta - A. \tag{4}$$

The first prism is placed in the position of minimum deviation for the mid-ray. Let the index of refraction for the latter be m. If we put m for n in (1), we must have:

$$\beta = \gamma = A/2; \quad \delta = \alpha.$$

These conditions determine the initial angle of incidence α (the same, of course, for all the rays). The total deflection of the mid-ray is $2\alpha - A$. The second prism is likewise placed in the position of minimum deviation for the mid-ray—but is oriented vertically, whereas the first prism is oriented horizontally (it should be recalled that the mid-ray has been assume to emerge horizontally from the first prism). Since all the rays that emerge from the first prism fall in a single vertical plane, it is clear that the angle p between any given ray falling on the second prism and the principal plane of that prism is the difference between the total deflection of that ray by the first prism, and that of the mid-ray—i.e., we have:

$$p = \Delta - 2\alpha + A. \tag{5}$$

(In expressing it so, we have chosen the upper half-space bounded by the horizontal principal plane as the "positive" one.)

For the second prism, let us designate the angles of incidence and refraction, and the total deflections, of the *projected* rays, by the same letters as were used for the actual rays in the case of the first prism, but distinguished by primes; and let A' be the refracting angle of the second prism. Since all the rays at their initial incidence upon the second prism have the same projection in its principal plane, the angle α' is the same for all; and since the second prism, like the first, is in the position of minimum deviation for the mid-ray, this angle is determined by conditions entirely analogous to those that determine α. For an arbitrary ray, then, in its passage through the second prism, we have:

$$\sin p = n\sin q; \tag{6}$$

$$n' \cos p = n \cos q; \tag{7}$$

$$\sin \alpha' = n' \sin \beta'; \tag{8}$$

$$\beta' + \gamma' = A'; \tag{9}$$

$$\sin \delta' = n' \sin \gamma'; \tag{10}$$

$$\Delta' = \alpha' + \delta' - A'. \tag{11}$$

In these equations, the quantities m, A, and A' are given; α and α' are thereby determined. All other quantities are to be regarded as functions of the parameter n: the index of refraction for (representing, in Newton's terms, the "refrangibility" of) the particular ray. To facilitate comparison, I shall use for the plane of the wall the same system of coordinates as Lohne: the origin is again to be the point reached by the mid-ray when only the first prism is present (or when the angle A' is zero); the x-axis is to be vertical, the y-axis horizontal (contrary to the usual custom). The perpendicular distance from the crossing-point of the prisms to the wall will again be represented by R. Then one easily sees that:

$$x = R \sec \Delta' \tan p; \tag{12}$$

$$y = R \tan \Delta'. \tag{13}$$

This system of equations does not lend itself to the kind of resolution into closed parametric form that Lohne achieves easily for his modified experiment. It is nevertheless not hard to derive from these equations quantitative results of some usefulness for our question. First, however, a qualitative comparison with Lohne's equations will be instructive.

The parameter φ in Lohne's equations corresponds, it should be recalled, to our Δ, and his quantity $\varphi - 37°11'$ corresponds to our p (the difference between the deflection of a given ray by the first prism and that of the mid-ray). To be sure, a qualification is required: Lohne's parameters and ours, although they *correspond* to one another, are not *identical*; for we must remember that Lohne rotates the first prism so as always to be in the position of minimum deviation for the ray concerned. However, a closer examination shows that this has negligible effect, so long as we confine our attention to a short spectral interval about the mid-ray. Indeed, our prism is in the position of minimum deviation for the mid-ray, and therefore near the position of minimum deviation for "nearby" rays; and (of course) the deviation (for a ray of given refrangibility) has a stationary value at minimum deviation. Let us therefore provisionally neglect the difference between Δ and φ (as functions ot the index of refraction n).

Now, the angles p and q, for rays in the optical range and for substances like glass or water, are quite small—of the order of 1° for glass, of 30′ for water (when the refracting angle of the first prism is 60°). The cosines of the angles p and q, therefore, are very close to 1, and their ratios are consequently also very close to 1: the ratio $\cos q/\cos p$, calculated for a prism of 60°, with Newton's values for the indices of refraction of his glass (namely, $81656/52400 = 1.558$ and $80481/52400 = 1.536$ for the extreme rays, $8107/5240 = 1.547$ for the mid-ray—see Newton [3], pp. 321, 333), is always less than 1.0001. Thus the effective index of refraction n' of a projected ray is indistinguishable from n, the true index of refraction of the (unprojected) ray; and therefore the projected deviation Δ' by the second prism is, for a given ray, practically the same as if it had been incident along the same line as the mid-ray. In the case represented by Lohne's equations, when the refracting angles of the two prisms are equal, it follows that Δ' is practically equal to Δ. Since the quantity $\cos(\varphi - 37°11')$—i.e., $\cos p$—in the denominator of Lohne's expression for y is practically 1, our equation (13) very nearly agrees with his in the visible range.

On the other hand, the factor $\sec \Delta'$ makes a significant difference between equation (12) and Lohne's expression for x. Its effect is that, whereas the horizontal projection of the line of the spectrum (given by equation (13)) is (nearly) the same as for Lohne's modified experiment, its vertical projection is—in comparison to his—appreciably stretched out. But this has the effect of increasing the slope of the line of the spectrum; and it follows that Newton's overestimate of that slope is smaller than Lohne implies.

The point is worth a little further exploration. Equation (12) shows that Newton's own diagram is in error in depicting the series of spectra—produced by a series of second prisms with different refracting angles—as having extremities that lie in two horizontal lines. Indeed, from equations (12) and (13) together we see that $x^2 \cot^2 p - y^2 = 1$; so, since the angle p is independent of the second prism, the line connecting the points on the several spectra corresponding to a ray of given refrangibility is a hyperbolic arc, convex towards the horizontal axis of our coordinate system. Now (somewhat ironically) on *this* point Lohne agrees with Newton: in his modified experiment, the lines in question are horizontal straight segments. The geometrical reason for this is very simple. The effect of Lohne's readjustment of the second prism for each ray is to make the plane of the second deflection always pass through the line of the first prism—a horizontal line, parallel to the wall. Thus the intersection of that plane with the wall is itself always horizontal. On the other hand, with the second prism always vertical, the angle p is independent of the second prism; therefore, for a given index of refraction, the twice-deflected ray falls always on a right circular cone with vertical axis—hence the hyperbolic intersection with the wall.

Proceeding now to quantitative estimates, it is pertinent to determine the slope of our spectral curve. From equations (12) and (13), we have for the slope at an arbitrary point:

$$\frac{dx}{dy} = \cos \Delta' \sec^2 p \frac{dp}{d\Delta'} + \sin \Delta' \tan p. \tag{14}$$

To evaluate $dp/d\Delta'$, we have to differentiate equations (1)–(11). This yields, in the first place, $dp = d\Delta = d\delta = n(\cos\gamma/\cos\delta)\, d\gamma + (sin\gamma/\cos\delta)\, dn$, from which we obtain:

$$dp = d\Delta = (\cos\gamma \tan\beta + \sin\gamma) \sec\delta \, dn. \tag{15}$$

Next, we observe that equations (8)–(11) exactly parallel equations (1)–(4); therefore:

$$d\Delta' = (\cos\gamma' \tan\beta' + \sin\gamma') \sec\delta' \, dn'. \tag{16}$$

Now, in the general case, if we calculate dn' in terms of dn and substitute in these equations, the result is rather cumbersome. But what is of principal interest is the slope of the spectral curve at the position of the mid-ray; for the optical part of the curve is a short arc, nearly coincident with a segment of its own tangent at that point. Differentiation of equation (7) yields: $\cos p \, dn' = \cos q \, dn - n \sin q \, dq + n' \sin p \, dp$. Since, for the mid-ray, $n = m$ and $q = p = 0$, we find for this case that $dn' = dn$. Again, for the mid-ray, $\alpha = \delta$, $\beta = \gamma = A/2$, $\alpha' = \delta'$, $\beta' = \gamma' = A'/2$, and $\Delta' = 2\alpha' - A'$. Taking account of these relations, and noting too that $\sin\beta = (1/m)\sin\alpha$, $\sin\beta' = (1/m)\sin\alpha'$, we see that $dp = (2/m)\tan\alpha \, dn$, $d\Delta' = (2/m)\tan\alpha' \, dn$, and, finally:

$$\left.\frac{dx}{dy}\right|_{n=m} = \cos(2\alpha' - A') \tan\alpha \cot\alpha'. \tag{17}$$

For the special case $A' = A$, the slope at $n = m$ is $\cos(2\alpha - A)$, i.e., $\cos\varphi$ (where now the identity of our deviation with Lohne's φ is rigorous). The analogous result for the slope of Lohne's curve at $n = m$ is easily seen to be $\cos^2\varphi$; and this confirms our qualitative conclusion that Lohne's slope is significantly smaller than the true one for Newton's experiment. If we use Newton's value $m = 1.547$ for the pair of 60° prisms, we find $\varphi = 41°20'$; the slope of Lohne's curve, then, is 0.564, whereas the true slope is 0.751. If Newton carried out the experiment not with prisms of glass, but with hollow water-filled prisms (a procedure he recommends when he describes this experiment in the *Opticks*—Newton [4], p. 78), then the deviation of the mid-ray becomes $23°54.3'$, and the slope is 0.833 for Lohne's curve, 0.913 for the true one.[22]

One further difference between the results of Lohne's modified experiment and the actual one is worth mentioning. If we continue to suppose that Newton used water-prisms, and calculate, for the case where the refracting angle of the second prism is 30° (that of the first being still 60°), the point of intersection of the vertical coordinate-axis with the tangent line to the spectrum at the position of the mid-ray, that point turns out to fall *below* the position of the undeviated solar image. This cannot happen according to Lohne's diagram, since his spectral curves are all convex towards the vertical axis.[23] The corresponding

[22]That Newton may have used prisms of water, and that this may partially account for his erroneous conclusion from this experiment, is suggested by Shapiro in Newton [3], p. 445, n. 4.

[23]As has already been noted, Lohne says, mistakenly, that these curves are circles. For the special case in which the prisms have equal refracting angles—the case to which Lohne's

intercepts for a pair of water-prisms with refracting angles both equal to 60°, and for glass prisms either both of 60° or the first 60° and the second 30°, all lie above the undeviated solar image (so that Newton's diagram is indeed incorrect in this respect); but in all three cases, Lohne's diagram remains qualitatively erroneous (for the actual experiment of Newton) in showing convexity towards the vertical axis. Each of these (true Newtonian) curves has a point of inflection for some value of n between 1 (which gives the undeviated image) and m, as one can see by calculating the slopes of the several curves at $n = 1$ (by a procedure analogous to that carried out above for $n = m$): it turns out for each of them that the slope at $n = 1$ is *less* than that at $n = m$.

References

Lohne, Johannes A.

[1] "Newton's 'Proof' of the Sine Law and his Mathematical Principles of Colors," *Archive for History of Exact Sciences,* **1** (1961), 299–405.

Neugebauer, Otto.

[1] *The Exact Sciences in Antiquity.* 2d ed., 1957; reprinted New York: Dover Publications, 1969.

[2] *A History of Ancient Mathematical Astronomy.* 3 vols. New York: Springer-Verlag, 1975.

Newton, Isaac.

[1] *The Correspondence of Isaac Newton.* Vol. 1, ed. H. W. Turnbull. Cambridge University Press, 1959.

[2] *Isaac Newton's Papers and Letters on Natural Philosophy.* Edited by I. Bernard Cohen. Cambridge, Mass: Harvard University Press, 1958.

[3] *The Optical Papers of Isaac Newton.* Vol. 1. *The Optical Lectures, 1670-1672,* ed. Alan E. Shapiro. Cambridge University Press, 1984.

---

equations apply, and in which his curve is the projected image of a hippopede—I have verified that his diagram is qualitatively correct in the relevant respect. More precisely: (a) for $0 < \varphi < 90°$, the curve is convex towards the vertical axis; (b) at $\varphi = 0$—where the curve, which is symmetric about the vertical axis, crosses that axis (with a self-intersection)—each branch has a point of inflection, and is thus, after crossing the axis, convex towards it once again. (Of course this last fact is *not* in accord with Lohne's conclusions; but his diagram does not show the extensions of his curves beyond their intersections with the vertical axis—extensions which, corresponding to anomalous dispersion with $n < 1$, have no bearing on Newton's experiments.) I have not checked to see whether these qualitative conditions continue to hold for Lohne's spectral curves in the more complicated case when the refracting angles of the two prisms are unequal.