VI. SUMMARY

I have argued that Newton's first two laws of motion are not definitions of force but rather contain strong statements about the nature of the physical world that are, in principle, falsifiable. In particular I argued that the first law postulated the existence of an ensemble of straight lines which, together with the Newtonian planes of absolute simultaneity defined the geometry of Newtonian mechanics. The transition to special relativity was effected by replacing these planes by light cones and the transition to general relativity by replacing the absolute geometry defined by these objects by a dynamical one.

ACKNOWLEDGMENT

I wish to thank James Supplee for several helpful discussions and suggestions on improving the presentation of this paper.

[1]J. B. Marion and S. T. Thornton, *Classical Dynamics of Particles and Systems* (Harcourt, San Diego, 1988), 3rd ed., p. 45.

[2]Sir A. S. Eddington, *The Nature of the Physical World* (Macmillan, New York, 1930).

[3]The formulation given here was first presented in J. L. Anderson, *Principles of Relativity Physics* (Academic, New York, 1966) and was further developed in M. Friedman, *Foundations of Space-Time Theories* (Princeton U. P., Princeton, 1983).

[4]The notion that the geometry of Newtonian space is that of a four-dimensional affine space appears to have originated with H. Weyl, *Space-Time-Matter*, translated by H. L. Bose (Dover, New York, 1922), p. 155. The mathematical formulation of this idea was first given by E. Cartan, "Sur les varietes a connexion affine et la theorie de la relativite generalisee (premiere partie)," Ann. Ecole Norm. Suppl. **40**, 325–412 (1923), "Sur les varietes a connexion affine et la theorie de la relativitie generalisee (suite)," Ann. Ecole Norm. Suppl. **41**, 1–25 (1924); see also, A. Trautman, "Sur la theorie newtonienne de la gravitation," C. R. (Paris) **257**, 617–620 (1963); J. L. Anderson, Ref. 3.

[5]J. L. Anderson, Ref. 3.

[6]This view was first expounded to me in a private discussion with Franklin Pollock.

[7]A. Trautman, Ref. 4.

[8]A. Einstein, "Zur electrodynamik bewegter Korper, "Ann. Phys. Leipzig **17**, 891–921 (1905).

[9]A. Einstein, "Zur allgemeinen Relativitatstheorie," Preuss. Akad. Wiss. Berlin, Sitzber. 778–786, 799–801 (1915); English translation available in A. Einstein, *The Principles of Relativity* (Dover, New York, 1952).

Precise calculation of the electrostatic force between charged spheres including induction effects

Jack A. Soules
Department of Physics, Cleveland State University, Cleveland, Ohio 44115

(Received 12 January 1989; accepted for publication 9 February 1990)

The method of images is applied iteratively to compute precisely the electrostatic force between charged spheres. The results are compared to the experimental results reported by Coulomb and show that Coulomb overlooked induction effects revealed in his data.

Most textbooks introduce electricity by describing Coulomb's experiment and his results which "can be represented by $F \sim 1/d^2$."[1] This has troubled me since I was a student. When the similarly charged balls are close enough to produce a detectable force, that force must be large enough to move the charges to the far sides of the spheres, thus producing a weaker force than $1/d^2$. In other words, the effective d has increased. When the balls are oppositely charged the attractive force brings the charges to the near sides so that the effective d is smaller than the center-to-center distance. Unwilling to settle for such a qualitative description, I decided to calculate the actual forces between charged spheres. The result is a tidy problem in mathematical physics suitable for undergraduates. Of course, the standard form of the force, $F \sim 1/d^2$, does apply to point charges that are incapable of polarization effects. It requires quantum electrodynamics to calculate those effects when two electrons are really close. After performing the calculations for the force between charged conducting spheres (see Appendix for details), we decided to compare our results to the actual experimental forces measured by Coulomb. Our calculated results (see Fig. 1) had confirmed our expectation that the repulsive force increases less rapidly than $1/d^2$ and that the attractive force increases much more rapidly.

A brief annotated translation of Coulomb's original paper[2] can be found in Magie.[3] Coulomb's description of his apparatus is somewhat vague. For example, the pith balls were "2 or 3 lines" ($\frac{1}{6}$ to $\frac{1}{4}$ in.) in diameter. The radius of the arm that supported the moving ball is not given but from the illustration it was evidently about 18 lines ($1\frac{1}{2}$ in.). (See Fig. 2.) The distance between the centers of the balls was

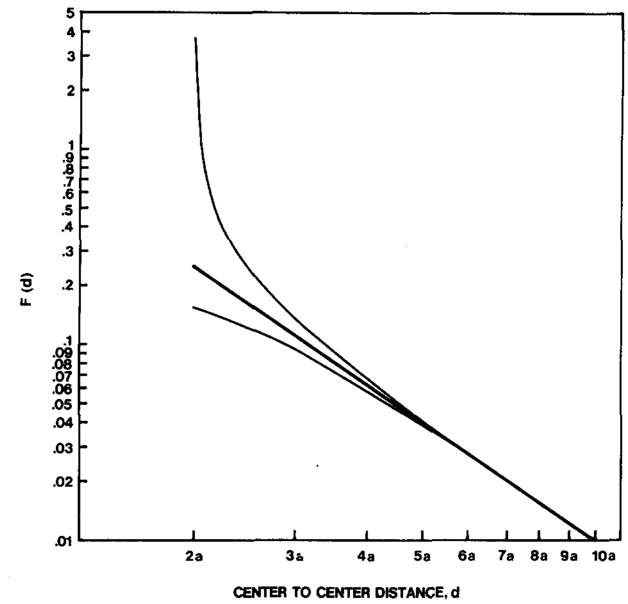

Fig. 1. Behavior of the electric force between identical charged spheres as a function of center-to-center distance.

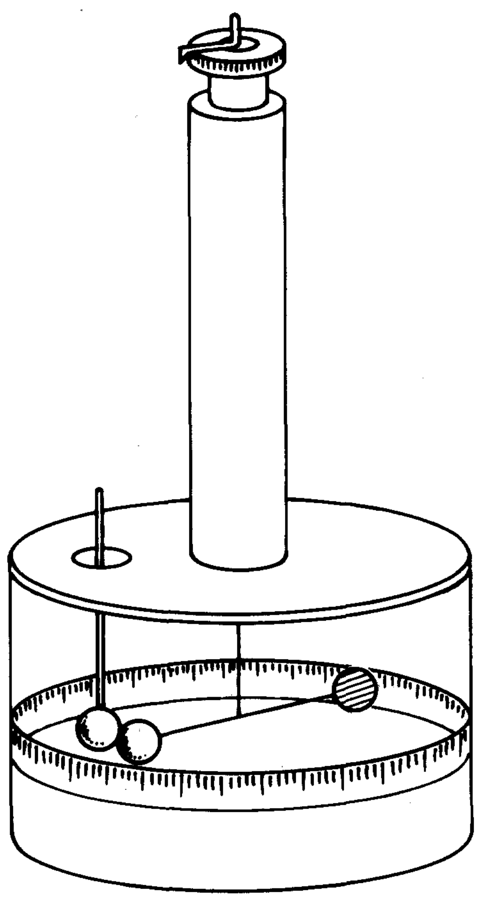

Fig. 2. Simplified diagram of Coulomb's apparatus for measuring electrostatic repulsion.

$d = 2a \sin \theta/2$, which Coulomb treated as approximately $a\theta$. He found that when the angle was decreased from 36° to 18°, the force quadrupled, within an error of less than 1%. However, to quadruple again the angle had to be reduced, not to 9°, but to $8\frac{1}{2}$°. Coulomb noticed this $\frac{1}{2}$° error but made no further comment on it. Evidently he did not expect his apparatus to give a more precise result even though it was capable of it. Of course, we know that the experiment was reporting the force accurately. It was the theory that was in error!

At 8° or 9°, the parameter d/a is somewhere between 3.7 and 5.5, where we see from Fig. 1 that the force is about 5% less than $1/d^2$. Coulomb was so eager to prove the $1/d^2$ law that he overlooked the experimental observation of polarization.[4]

Because Coulomb did not specify the exact diameters of his pith balls, we also calculated the repulsive force curve for balls with diameters in the ratio 3:2. The resulting force at the distance $d/(a_1 + a_2) = 2$ or more is only slightly less ($\sim 1\%$) than the force shown in Fig. 1, leaving our conclusion intact. However, at close approach, the repulsion between dissimilar balls falls significantly ($\sim 50\%$) below the curve for identical balls shown in Fig. 1.

It is too much to expect Coulomb to have understood the concept of equipotential surfaces in 1785. However, he did know that the electrical "fluid" was mobile on the pith balls since if it were not so he could not have charged them. He failed to recognize, however, that this same mobility would cause the charges to move on the balls in the presence of the electrical force, thus spoiling his effort to identify the center of a ball with the center of the charge.

Coulomb's torsion balance was and is[5] an extremely sensitive and precise device for physical measurements. However, it is a small miscarriage of justice to name the electrostatic force law in his honor. Most scientists of his time expected the force law to be $1/d^2$ and others[6] had measured it to be $1/(d^{2+\epsilon})$, for example. It was Priestley[7] who recognized that if cork balls inside a charged cup were unaffected by the charge on the cup then $F \sim 1/d^2$ exactly. He drew this conclusion from Newton's work on the force of gravity within a hollow sphere. This theoretical insight is far more satisfactory in establishing $F \sim 1/d^2$ than *any* experiment, which, as can be worked out from Table I, will give a force law more like $F_{\text{rep}} \sim 1/D^2 - 1/D^4 - 2/D^6$, etc., where $D = d/a$.

When Coulomb attempted to measure the attractive force between unlike charges he met a serious problem. The force increases so rapidly at small distances that his torsion balance, for which the force was linear with distance, could not compensate. To measure the attractive force he used a suspension with so little restoring torque that its natural period of oscillation was very long. He then brought a large charged sphere close to a suspended disk and measured the period under conditions in which the restoring force was, for all practical purposes, the electrical force. (See Fig. 3.) Coulomb's large sphere in this experiment was 6 in. in radius and the small disk was about $\frac{7}{24}$ in. in radius. The separation in various trials ranged from 9 in. (center-to-center) to 18 in. Coulomb tells us he charged the disk by induction when it was "some inches" away.

There are two ways of estimating the charge on the small disk. A disk carrying charge q has a potential $1.5708 (= \pi/2)$ times the potential of a sphere of the same radius.[8] If we compute the potential at the center of the disk

Table I. Calculated values of F as a function of center-to-center distance.

Distance D as a multiple of a	F attraction	F Coulomb	F repulsion	F^{*a}
10.0	0.010 041	0.01	0.009 96	0.009 990
9.0	0.012 415	0.012 346	0.012 277	0.012 1190
8.0	0.015 751	0.015 625	0.015 501	0.015 373
7.0	0.020 656	0.020 408	0.020 165	0.019 975
6.0	0.028 324	0.027 778	0.027 250	0.026 963
5.0	0.041 404	0.04	0.036 680	0.038 272
4.0	0.067 097	0.062 5	0.058 457	0.058 106
3.0	0.134 819	0.111 111	0.094 437	0.096 022
2.5	0.240 647	0.16	0.121 091	0.126 208
2.4	0.285 277	0.173 6	0.127 089	0.133 005
2.3	0.352 514	0.189 036	0.133 313	0.139 797
2.2	0.470 269	0.206 612	0.139 794	0.146 283
2.1	0.759 739	0.226 757	0.146 583	0.152 019
2.05	1.225 38	0.237 954	0.150 12	0.154 385
2.01	3.764 5	0.247 519	0.154	0.155 924
2.0	∞	0.25	0.154	0.156 25

[a] $F^* = 1/D^2 - 1/D^4 - 2/D^6$ with $D = d/a$; F^* is a simple function that closely approximated $F_{\text{repulsion}}$.

due to the large sphere and place a charge on the disk of opposite sign so as to reduce its potential to zero, we get $q_{\text{disk}} = 0.014\,920_{\text{sphere}}$ when the disk is 8 in. from the center of the large sphere. This calculation ignores induction effects on the large sphere. Alternatively we can replace the disk with a small sphere of equal capacitance, in which case the small sphere will have a radius of $2/\pi \times \frac{7}{144} \times a_{\text{large}} = 0.0309 a_{\text{large}}$. Using the image charge iteration technique described in the Appendix, we find that the small sphere is at zero potential when it is charged to about -0.0240_{sphere}, about 50% more charge!

With these figures in hand, we can proceed to calculate what Coulomb should have observed for the attractive force. Of course, the actual charge on the disk could be considerably different from our estimated value since we have no way of knowing what Coulomb meant by "some inches." In the absence of induction effects, the charge on the small disk would not matter in determining the force law. However, in the presence of induction it clearly does.

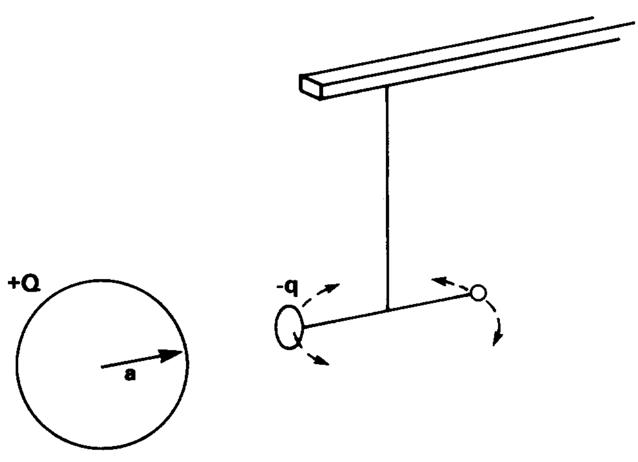

Fig. 3. Simplified diagram of Coulomb's apparatus for measuring electrostatic attraction.

We know that even in the event $q_{\text{disk}} = 0$ there is an attractive force! We return then to our iterative procedure to compute the attractive force between a large sphere ($a_{\text{large}} = 1$) and a small sphere ($a_{\text{small}} = \frac{1}{32}$) carrying Q_0 and $-0.0240 Q_0$, respectively.

Table II compares the results of the force computation including induction with the pure $1/d^2$ results. We have suppressed the $1/4\pi\epsilon_0$ factor and the ratio of inches to meters in both columns. It is clear that the attractive force is much stronger than $1/d^2$ for $d/a < 1.5$.

Coulomb's torsional oscillator should have a period inversely proportional to $F^{1/2}$, therefore, directly proportional to d if $F \sim 1/d^2$. He presents his results in the form of a table:

The distances are as the numbers: 3 6 8.

The times of the same numbers of oscillations are: 20 41 60.

By theory they ought to have been: 20 40 54.

He then explains the small discrepancy at large distance by claiming that the charge had probably drained away a bit. (Note: He evidently did not reverse the order of the experiment to see if that was *really* the case!)

From our enlightened position, we choose to present his data differently, making the period at large separation the most reliable. We then have the following.

The distances are as the numbers: 8(24 in.) 6(18 in.) 3(9 in.).

By theory ($1/d^2$) the times should be: 60 45 22.5.

By theory (corrected for induction) the time should be 60 45 21.5.

By experiment the times are: 60 41 20.

Clearly, induction effects can account for a substantial part of the discrepancy.

The intellectual climate of the 18th century was such that Coulomb could not hope to gain much respect from inventing a clever and sensitive apparatus for measuring forces. Proving that electrical charges obeyed a force law identical to the gravitational force law, on the other hand, would have great philosophical significance. We should not be surprised, therefore, that Coulomb "stretched" his data to "prove" the $1/d^2$ law rather than to respect the precision of his instrument enough to find an entirely new electrical effect.

Since our calculation technique was so convenient we proceeded to examine the induced force in two additional cases: (1) when an uncharged sphere is brought near to a charged one; and (2) when two equally charged spheres are of different radii.

Table II. Representative forces in simulation of Coulomb's attractive experiment.

Distance D as a multiple of a	Force with induction	Force, $1/D^2$
3.0	0.111	0.111
2.5	0.161	0.160
2.0	0.252	0.250
1.5	0.462	0.446
1.333	0.604	0.562
1.167	0.942	0.733
1.083	1.812	0.850

When our very small sphere was brought, uncharged, near to the large sphere the resulting force which we calculated was essentially zero. One can interpret that result either by saying that the gradient of the field is too small over the diameter of the small sphere or by noting that the image charges are small and highly localized on both spheres. When the small sphere is charged, however, and brought near the uncharged large sphere the resulting force is very large. Essentially, the small sphere sees its negative image close by and is strongly attracted to it. Or, in a different description, the potential of the small sphere is large and the field gradient over the diameter of the large sphere is also large.

When the two spheres are of different radii a surprising result occurs. If $a_1 > 1.24a_2$ there exists a center-to-center spacing where the force between equally charged spheres becomes attractive. Of course, in the laboratory we seldom deal with equal charges on dissimilar spheres. It is more common to charge the spheres to same potential.

In Fig. 4 is shown the position of the null point, at which the force between equally charged spheres changes sign from repulsion to attraction. Distances in Fig. 4 are expressed as $d/(a_1 + a_2)$ and the abscissa is the "eccentricity" or a_1/a_2.

These results should not surprise us. If the larger sphere were very much larger than the smaller one its charge would create only a small repulsive force on the small sphere. The small sphere, on the other hand, would be attracted to its own oppositely charged image by a much larger force. Our calculations show that the onset of this effect occurs at about $a_1 = 1.24a_2$.

The program we have written is convenient for the solution of any electrostatic problem involving two spheres, without the need for the solution of a differential equation with boundary conditions.

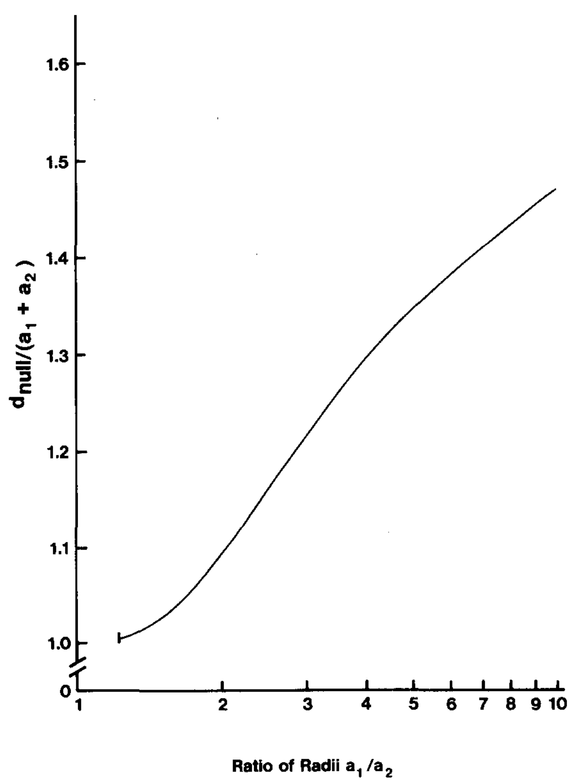

Fig. 4. Unequal spheres, equally charged; center-to-center distance where force changes sign.

APPENDIX

Consider two similar charges, q_0, a distance d apart. If we surround one charge with a conducting sphere of radius a and require that the sphere be an equipotential surface, we find that an image charge of magnitude

$$q_1 = -aq_0/d \tag{A1}$$

must be placed a distance a^2/d from the center of the sphere to produce a spherical equipotential at the radius a.[2] The original charge q_0 at the center of the sphere must be increased to $q_0^+ |q_1|$ to maintain the total charge equal to q_0. A second conducting sphere of radius a around the second charge q_0 will be an equipotential if two new image charges are placed inside, corresponding to the images of q_1 and of $q_0^+ |q_1|$. Each image charge is smaller than its object charge and of opposite sign. By repeated iteration of this scheme, a task well suited to a personal computer, we found that a string of charges of alternating sign and rapidly decreasing magnitude can replace the charged spheres and produce the appropriate field. Figure 5 shows the locations of the image charges when $d = 2.1\,a$.

The computer program calculates all of the q_n and x_n from

$$D_n = d - x_{n+1},$$

$$x_{n+1} = 1/D_n,$$

and

$$q_{n+1} = -q_n/D_n, \tag{A2}$$

with

$$q_0 = 1,$$

$$D_0 = d,$$

and

$$x_0 = 0.$$

The resulting charges are in the right proportions but no longer sum up to Q_0 so their sum S is calculated and a new $q_0 = 1/S$ is chosen. The process is repeated until $|S - 1| < 10^{-6}$. To calculate the total force between the spheres we simply sum all of the forces between the sets of charges, a task that the computer does very quickly. In practice, when the distance between the spheres d is greater than $4a$, as few as 10 charges suffice for six-digit accuracy. On the other hand, when the spheres touch, $d = 2a$, and more than 100 charges on each side are needed to compute

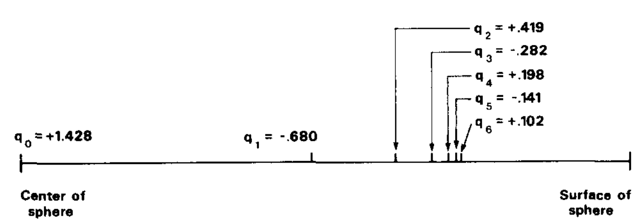

Fig. 5. Distribution of image charges when $D = 2.1a$, repulsive case.

the force. Because the charges alternate in sign, the series does converge for all $d > 2a$.

The problem of the attractive force between unlike charges is tackled in the same way. The charges q_i' appear at the same positions and with the same magnitudes but this time are all of one sign, negative on one side and positive on the other. The resulting total force between the charge systems is slightly larger than a^2/d^2 but, unlike the repulsive case, begins to diverge rapidly as $d \to 2a$. In Fig. 1 are plotted the logarithms of the repulsive, the a^2/d^2, and the attractive forces as a function of the logarithm of the parameter d/a. Of course, the a^2/d^2 graph is a straight line. The distribution of image charges in a typical example is shown in Fig. 5.

The calculation of electric potentials, fields, and charge distributions by the method of images is described in many undergraduate and graduate electricity texts. Smythe's text[8] is particularly useful because it was written after the appearance of the computer, which has made formerly impossibly tedious calculations feasible. Smythe, for example, shows that the image charges for two conducting spheres satisfy a difference equation that can be solved in closed form:

$$q_n = R_1 R_2 \sinh \alpha / R_2 \sinh n\alpha + R_1 \sinh (n-1)\alpha$$

with

$$\alpha = \cosh^{-1}(d^2 - R_1^2 - R_2^2/2R_1R_2).$$

We have chosen to present our calculations in computer iterative form simply because the interesting questions of physics and of mathematics (convergence, divergence) are more transparent in that form. Either method can be used to calculate the potentials, forces, and capacitances between conducting spheres of arbitrary radius and spacing.

[1] David Halliday and Robert Resnick, *Fundamentals of Physics* (Wiley, New York, 1986), 2nd ed., p. 457. Most elementary texts have a similar statement.

[2] Charles-Augustin Coulomb, "Law of electric force," Mem. Acad. R. Sci. 569, 578, (1785).

[3] William Francis Magie, *A Source Book in Physics* (McGraw-Hill, New York, 1935), p. 408.

[4] Coulomb's many papers were reprinted in 1884 by Gauthier–Villars under the auspices of the Societé Francaise de Physique. This is the version that is most commonly available in American libraries. In the reprinted version of the 1785 papers on the electrostatic force there are several insertions in which the various assumptions of Coulomb are analyzed mathematically. In the first of these, the annotator calls attention to the small error due to Coulomb's failure to account for inductive effects. Unfortunately, the annotator is not otherwise identified.

[5] Gabriel G. Luther and William R. Towler, "Redetermination of the Newtonian gravitational constant G," Phys. Rev. Lett. **48**, 121–123 (1982).

[6] Cf., for example, the discussion in Franklin Miller, *College Physics* (Harcourt Brace Jovanovich, New York, 1972), 3rd ed., p. 372.

[7] Reference 6.

[8] Cf. Wolfgang K. H. Panofsky and Melba Phillips, *Classical Electricity and Magnetism* (Addison-Wesley, Reading, MA, 1955), p. 85 and Edmund T. Whittaker and George N. Watson, *A Course of Modern Analysis* (Cambridge U. P., London, 1963), p. 382. Panofsky and Phillips show that the potential of the disk is

$$V(a) = \frac{q}{4}\pi\epsilon_0 a \int_0^\infty x^{-1} J_0(x) \sin x \, dx.$$

Whittaker and Watson show that $\int_0^\infty x^{-1} J_0(xt) \sin x \, dx = \pi/2$, for $0 < t < 1$.

[9] William R. Smythe, *Static and Dynamic Electricity* (McGraw-Hill, New York, 1968), 3rd ed., pp. 128–131.