



Robert A. Millikan meets the credibility revolution: comment on Harrison (2013), 'field experiments and methodological intolerance'

Nathaniel T. Wilcox

To cite this article: Nathaniel T. Wilcox (2016): Robert A. Millikan meets the credibility revolution: comment on Harrison (2013), 'field experiments and methodological intolerance', Journal of Economic Methodology

To link to this article: <http://dx.doi.org/10.1080/1350178X.2016.1158950>



Published online: 29 Mar 2016.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

Robert A. Millikan meets the credibility revolution: comment on Harrison (2013), ‘field experiments and methodological intolerance’¹

Nathaniel T. Wilcox*

Economic Science Institute, Chapman University, Orange, CA, USA

Millikan’s famous oil drop experiment is scrutinized from the viewpoint of the methodological dicta of the credibility revolution.

Keywords: Millikan; oil drop experiment; credibility revolution

Glenn Harrison’s (2013) initial *reductio* – that laboratory classics (e.g., Smith, 1962) involving no crucial randomization but plenty of control are somehow not experiments under the dictates of the credibility revolution – has to make any sane scholar laugh. But are there really people who believe things like that? You bet. In nearly the same breath, Harrison quotes Gerber and Green (2012, p. 9 n. 7) as follows:

Readers with a background in the natural sciences may find it surprising that random assignment is an integral part of the definition of a social science experiment. Why is random assignment often unnecessary in experiments in, for example, physics? Part of the answer is that the ‘subjects’ in these experiments – e.g. electrons – are more or less interchangeable, and so the method used to assign subjects to treatment is inconsequential. Another part of the answer is that lab conditions neutralize all forces other than the treatment.

Harrison judges this is ‘Nonsense, at several levels, but not worth dissecting since readers in social sciences are not expected to be puzzled by this’. Perhaps, yet this quote from Gerber and Green made me wonder: Is great empirical practice in the natural sciences so very different from our own? Is it better or worse? Would we recognize its ways as our own, or not? Would the prophets of the so-called credibility revolution of empirical economics approve of those ways, or not? If not, would or should natural scientists care much?

In high school physics, my classmates and I tried our hands at balancing a tiny charged drop of latex in the air between two electrified plates, mostly convincing us that our teacher had a cruel sense of humor. Otherwise I forgot all about Millikan’s famous ‘oil drop’ experiments until I read Gerber and Green’s assurances about the natural sciences. Curiosity took me back to Millikan’s (1911) first full published report (hereafter M11) on his attempt to (a) prove that electric charge is quantized and (b) measure that elementary electrical charge. Franklin (1981, p. 191) believes that ‘Millikan regarded the question of charge quantization as settled’ by M11, but there were still lingering criticisms of that conclusion – as well as Millikan’s measurement of

*Email: nwilcox@chapman.edu

¹I’m in debt to Nancy Cartwright, Allan Franklin and Christopher Ruf for their help navigating philosophical, physical and historical matters, but the usual disclaimer applies strongly – they’re not responsible for my bloopers.

the elementary charge. Millikan (1913) duly reported improved measurements and some answers to critics (hereafter M13). Millikan won a Nobel prize for this work, so perhaps it contains some of the essence of great experimental design and practice.

Here is the statistical essence of Millikan's deceptively simple problem. We have the linear regression of a dependent variable Y_i , measured with errors ε_i , on an independent non-negative variable X_i . This is the linear regression equation $Y_i = b + mX_i + \varepsilon_i$. Ultimately, Millikan wants the most precise estimate \hat{b} of the intercept b he can get (according to standard physical theory, this intercept will be a known monotonic transformation of the elementary charge Millikan seeks). Any regression text gives \hat{S}_b , an estimator of the standard error of \hat{b} , as

$$\hat{S}_b = N^{-1/2} \hat{S}_\varepsilon P, \quad (1)$$

where $\hat{S}_\varepsilon = \sqrt{\frac{\sum_{i=1}^N \hat{\varepsilon}_i^2}{N-2}}$ (with $\hat{\varepsilon}_i = Y_i - \hat{b} - \hat{m}X_i$), $P = \sqrt{\frac{SS(X_i)}{SS(X_i - \bar{X})}}$, $\bar{X} = N^{-1} \sum_{i=1}^N X_i$, $SS(X_i) = \sum_{i=1}^N X_i^2$, $SS(X_i - \bar{X}) = \sum_{i=1}^N (X_i - \bar{X})^2$, and

\hat{b} and \hat{m} are the OLS estimators of b and m .

How can Millikan make \hat{S}_b small by experimental design and method? Aside from increasing the sample size N , one avenue is to make \hat{S}_ε , the root-mean-square error or RMSE, as small as possible, and Millikan puts many different efforts here. The physical, mathematical path from Millikan's observations (the measured speeds of his oil drops) to the dependent variable Y_i passes through functional forms containing several physical parameters that needed their own measurement, either by Millikan's team or other experimenters elsewhere (more on this shortly). Insuring the precision of these other measurements was a central part of Millikan's efforts to reduce the measurement error variance in Y_i . We can call those other measurements covariates, since Millikan needs to control for them in order to properly interpret his dependent speed measures and construct his dependent variable Y_i from his primary oil drop speed measurements.

The other major term in the expression for \hat{S}_b is $P = \sqrt{SS(X_i)/SS(X_i - \bar{X})}$, the root of the ratio of the uncentered and centered sums of squares of the independent variable X . There is seemingly obvious advice here: Choose all the X_i equal to zero, reducing P to its minimum possible value of 1, and make \hat{S}_b depend solely on the RMSE \hat{S}_ε . Intuitively, Robert should just run observations with $X_i = 0$ and nix the regression: He is only after the intercept after all, the mean value of Y conditional on $X = 0$. At first blush, there appears to be no need to estimate the treatment effect of variations in X on Y . But the physics, and the limitations of the observational technologies of Millikan's day, preclude this: Observations of Y_i will have impossibly high measurement error variance for X near or at zero, or so Millikan repeatedly argues and tries to illustrate with the help of his experiences with relatively small but nonzero values of X . Shockingly, the inference Millikan wants to draw concerns the conditional expectation of Y at a value of X that cannot be within his sample range of X : Millikan is in the business of doing the impossible (or at least the statistically doubtful). The central thing of physical interest, the intercept b , is unavoidably an out-of-sample quantity – given the observational limitations of Millikan's day. It will have to be projected on the basis of estimated in-sample variations in $E(Y|X)$ and enough (assumed) local linearity. Indeed, we can view the term P in (1) as a 'penalty' exacted since the intercept cannot actually be directly measured with observations at $X = 0$.

The physics must eventually intrude on these mundane statistical musings. Once Millikan brings an oil drop between his electric plates, he measures its speed under two conditions. First, there is v_1 , the terminal speed at which drops fall under the force of gravity (balanced by the buoyancy and retarding force of the air) with the plates turned off; and there is v_2 , the speed at which drops rise under the added electromagnetic force with the plates on (due to the total charge a drop carries during that measurement). The drops themselves are thus Millikan's observational units or 'subjects': The subject may be electrons, but the subjects are oil drops.

Millikan repeatedly measures v_1 and v_2 for each drop, alternating between having the plates off or on:

This operation is repeated and the speeds checked an indefinite number of times, or until the droplet catches an ion from among those which exist normally in the air ... The fact that an ion has been caught ... is signaled to the observer by the change in the speed of the droplet under the influence of the field. (M11, p. 353)

So changes in total charge (brought about by naturally occurring ions) occur naturally in a wholly uncontrolled way, and Millikan might have carried off his measurement using only these uncontrolled variations in charge. However, Millikan does have a way to induce *partially* controlled changes in total charge as well, by bringing a tube of radium near the apparatus, and does so frequently: Millikan remarks that

This sort of a change was one which, after the phenomenon had once been got under control, we could make at will in either direction; i.e. we could force charges of either sign or in any desired number, *within limits* [my emphasis], upon a given drop. (M11, p. 366)

The contrast between v_1 and v_2 measures the *apparent* total charge e_n on a drop, which is further thought of as the product of the *apparent* elementary charge e_1 and the total number n of charges (electrons) carried. Here is Millikan's first pass at a formula (M11 p. 354 Equation 4) – elaborated later to account for a crucial nuisance (and this is why I say *apparent* charge at this stage of the analysis):

$$e_n = \frac{4}{3} \pi \left(\frac{9\mu}{2} \right)^{1.5} [g(\sigma - \rho)]^{-0.5} \frac{(v_1 + v_2)v_1^{0.5}}{\mathcal{F}}. \quad (2)$$

The actual procedure involves multiple rounds of measurement and computation for each given drop. At the start of a new round, the charge on a given drop is changed using the radium, a new measurement v_2 is taken and a new value of e_n computed. Doing this repeatedly, and appealing to the maintained hypothesis that e_1 is constant across rounds for a given drop, one may recover a maximum divisor e_1 of the several e_n figures that produces integer estimates of n (total electrons) carried by the drop for all that drop's measurement rounds. This e_1 is the *apparent* elementary charge as determined by the behavior of a given drop: It is a 'within-drop estimator' of the apparent elementary charge. It will need correcting for that crucial nuisance and we return to this later: The 'between-drops variation' of these within-drop estimates will be central to making this correction.

To use (2) to infer total charge from speeds, physical theory at Millikan's time demands knowledge of several other quantities. These are as follows:

- μ = the coefficient of viscosity of the air in Millikan's apparatus;
- g = the gravitational force in Millikan's apparatus (Franklin, 1981, p. 187);
- σ = the density of an oil drop in Millikan's apparatus;

ρ = the density of the air in Millikan's apparatus; and

\mathcal{F} = the strength of the electrical field surrounding the drop in Millikan's apparatus.

Initially, I thought most of these would be among those mythical context-invariant constants of the universe I have heard about, measured accurately with vanishing error. But this was not to be. Three of these – μ , σ , and ρ – vary with temperature, and Ryerson Hall was drafty in Millikan's day: In the first experiment, Millikan needed careful air temperature records and these varied from 19.9 to 27.8 C across the distinct drops used for calculations (M11, p. 384, Table 14). Millikan successfully controlled temperature to a much smaller range of 1.1 C around 23 C in his second experiment, but still took account of that smaller variation when making calculations (M13, pp. 134–135, Table 20).

Millikan has strong concerns about sources of variance in μ (the viscosity of air) as well as its mean at given temperatures. He says his measurement of the elementary charge will be 'limited in accuracy only by that attainable in the measurement of the coefficient of viscosity of air (M11, p. 350)' and that 'This factor certainly introduces as large an element of uncertainty as inheres anywhere in the oil-drop method (M13, p. 112)'. Given prior knowledge available at his time, Millikan chose this formula for estimating μ_T , the viscosity of air at temperature T :

$$\mu_T = \mu_{T'} + 0.000000493 \cdot (T - T') \quad (3)$$

where

T = temperature (deg C.) in the apparatus;

T' = a reference temperature (deg C.); and

$\mu_{T'}$ = a prior estimate of the viscosity of air at the reference temperature T' .

To put (3) to work, Millikan needs to select or create a reasonable prior estimate of $\mu_{T'}$ along with a reasonable guess about the 'likely error' in it (its irreducible uncertainty at his time). Millikan settles on a value and likely error of μ_{15} for the first study, depending on others' estimates and his own judgment as to which estimates are most reliable (M11, pp. 385–387). For the second study, a different value and likely error of μ_{23} , based on two measurements by Millikan's team in their own laboratory, was used (M13, pp. 112–115).

Millikan used an array of power cells to create field strengths \mathcal{F} ranging from about 8000 to 3000 v cm⁻¹ in M11 (p. 352), with \mathcal{F} gradually falling as an experimental session lengthened and the power cells discharged. In the second experiment, Millikan arranged for much closer control of field strength, employing a 5300-volt storage battery with a vastly slower discharge rate of 5–10 v h⁻¹, but field strength still varied across the 58 drops from about 3250–1050 v cm⁻¹ (M13, pp. 134–135, Table 20) and still required frequent measurement for subsequent calculations.

I have not touched on the calibration of instruments, but still we can enjoy all this work aimed at the reduction in measurement error, the resulting smaller \hat{S}_e , and finally the prize – the resulting reduction in \hat{S}_b . But now I must elaborate on the penalty P – why it exists at all, and things Millikan did to reduce it. The basic trouble stems from the behavior of relatively small oil drops. Equation (2) comes from two basic physical laws, Newton's second law (force is mass times acceleration) and Stokes' law. Stokes' law describes the 'retarding force' with which a *continuous* medium of viscosity μ resists a sphere of radius a moving through it, such as an oil drop moving through the air. The trouble comes from drops of sufficiently small radius. Being a mixture of molecules, air is not really continuous. At any given air pressure p , smaller oil drops are increasingly likely to slip between air molecules, dropping the retarding force of

the air below that specified by Stokes' law which, after all, is meant for a continuous medium. Because of this, Equation (2) is not exactly right and is increasingly wrong for smaller drops and/or at lower air pressures. We can summarize this by the product of the drop radius and the air pressure and say that Equation (2) is most accurate when ap is relatively large and increasingly misleading as ap approaches zero. In Equation (1), the independent variable X is simply the reciprocal of ap . So as ap gets large, X goes to zero, Stokes' law holds more perfectly, and we might measure the intercept in Equation (1) directly rather than by an out-of-sample projection. Why cannot Millikan arrange observation at sufficiently large ap ?

Under the forces of gravity and viscosity of air alone, faster drops are bigger drops. The force of gravity is proportional to drop mass, but the retarding force of viscosity is proportional to drop radius – and drop radius is less than proportional to drop mass. Therefore, as a drop's mass (and hence its radius) increases, so too its terminal velocity increases. Air pressure held constant, then, a relatively large ap means a relatively fast-moving drop. Keep in mind that the technology for measuring drop speed is very human: The experimenter watches for the time a drop takes to traverse a pre-measured short distance, using a chronograph or a stopwatch. The speed of fast-moving drops will be particularly difficult to measure accurately by this technique. Millikan (M11, p. 382) says that 'When the velocities are ... very large the time determination becomes unreliable ...' Throughout both papers, the point recurs frequently: '... it was thought preferable to make a considerable number of observations ... on drops small enough to make [v_1] determinable with great accuracy ...' (M13, p. 139).

Millikan is convinced that a fast drop, which is a big drop, is a bad drop from the perspective of keeping \hat{S}_b small. So much for choosing X close to zero. But Millikan is also convinced that a slow drop, which is a small drop, is also a bad drop. We have already seen that Stokes' law breaks down for sufficiently small drops, but there is another problem with the smallest ones: 'When the velocities are very small residual convection currents [in the air] and Brownian movements introduce errors' (M11, p. 382). So there is a Goldilocks range of just right speeds that Millikan decides to give particular credence. Although a handful of slower and faster drops will make the final cut,

[F]or the end here sought, namely, the most accurate possible determination of e , it was found desirable to keep the [inverse of the speed v_1] for the most part between 10 s and 40 s, in order to avoid chronograph errors on the one hand and Brownian movement irregularities on the other. (M13, p. 133)

I wish we knew how ten and forty seconds got privileged in this way, but we do not. All we really know is that there are good qualitative reasons for regarding the slowest and fastest drop measurements with suspicion. But make no mistake: Millikan is, with this (somewhat fuzzy) range requirement, selecting his subjects on the basis of values of their observed dependent measures. I see fingers wagging and scandalized Victorian grandmother looks on certain faces, yet this gross violation of alleged rules of method is closely reasoned and eminently reasonable. This selection issue is particularly amusing since we also know that Millikan selected and rejected many other drops for much less apparent reasons (Franklin, 1981; Holton, 1978). But after nearly 40 years of tut-tutting over *that other* selection of drops by Millikan, this much plainer selection – right there for all to see in the published work – has attracted no similar censure, though a sufficiently pedantic second-year econometrics grad student would happily give you a gratuitous earful about selection on values of the dependent variable.

To recap, considerations on likely sources of measurement error imply that the range of drop speeds that will get used in the analysis, and hence the drop radii a that will get used, must be limited to some range. This means that X in Equation (1), the reciprocal of ap , will have a limited range as well. To my mind, Millikan's greatest design innovation in passing from the 1911 article to the 1913 article was seeing how he could make experimentally controlled variations in air pressure p a huge advantage in his effort to reduce \hat{S}_b . In the 1911 experiment, uncontrolled variations in air pressure were simply nuisance variance to be measured and adjusted for in calculations, and the vast majority of variation in ap available for estimation came from the naturally occurring heterogeneity of drop sizes. In the 1913 experiment, controlled variations in air pressure would instead be an ally, since these allowed Millikan to vastly extend the range of the independent variable $X = (ap)^{-1}$ beyond the speed (and hence radius) limits Millikan imposed on himself to limit measurement error. Extending the range of X in this way permits very pronounced reductions in the penalty term $P = \sqrt{SS(X_i)/SS(X_i - \bar{X})}$ that appears in the expression for \hat{S}_b in Equation (1).

Stokes' law says that the retarding force on a sphere of radius a traveling through a continuous medium of viscosity μ at speed v is $6\pi\mu av$. Millikan's fix-up for the non-continuous nature of air is simply to discount (divide) this force by a factor $1 + c(ap)^{-1}$, a bald linearization of whatever functional form the breakdown in Stokes' law might take, given flesh with a single unknown parameter c (M13, pp. 110–111, 133–138). With this done, we have this relationship between the *actual* elementary charge e and the *apparent* elementary charges e_1 as computed for the drops using Equation (2) as described earlier:

$$e_1^{2/3} = e^{2/3} \left[1 + c(ap)^{-1} \right] = e^{2/3} + ce^{2/3}(ap)^{-1} \quad (4)$$

With Equation (4), we now have the simple linear regression of Equation (1). The dependent variable is $Y_i = e_1^{2/3}$, the independent variable is $X_i = (ap)^{-1}$, the slope parameter is $m = ce^{2/3}$ and, as promised, the intercept $b = e^{2/3}$ is a known monotonic transformation of the elusive elementary charge.

What have I learned about Millikan's methods? Consider the subjects – the oil drops. First, they are most definitely *not* interchangeable: *Subjects are heterogeneous in both controlled and uncontrolled ways.* The drops come in different sizes, carrying different and only partially controlled charges, with the charges changing (within subject) from round to round and observed under deliberately varied (between subject) conditions of air pressure (in M13 only). Second, *the heterogeneity of the subjects – both controlled and uncontrolled – is crucially central to the identification strategy.* Because it is not possible to accurately measure the behavior (terminal speed) of the theoretically ideal oil drop (e.g., one the size of a basketball at a hundred atmospheres) as it traverses the centimeter or so between the electrical plates, an out-of-sample projection will need to be made; and sample variation in the drops' values of $(ap)^{-1}$ is crucially central to this projection. Nor is that all: Third, *heterogeneity of subjects is crucially central to making the standard error of the central estimate small.* In Millikan's first experiment, the naturally occurring variation in drop radii is almost the *entire* variation available for estimating the central correction to Stokes' law. In his second experiment, Millikan deliberately uses air pressure variations to increase the range of the crucial covariate $(ap)^{-1}$ so as to reduce the penalty term P in Equation (1). This may all come as a surprise to acolytes of the credibility revolution: Among them, heterogeneity only plays an infernal role in a Manichean story about the damned nuisance it can be as seekers follow the one true path to unbiasedness.

This is as good a place as any to notice how obsessed Millikan is with minimum variance estimation rather than unbiasedness. You would have to read him very carefully to miss this. I have said nothing at all about calibration of instruments, but many pages of the two papers are on nothing else. I have mentioned the search for the very best parameter estimates for the interpretation and use of temperature (and viscosity and density) covariates. The beautiful use of air pressure variation to extend the range of $(ap)^{-1}$ has everything to do with reducing the central standard error by means of design artifice and nothing at all to do with unbiasedness.

The prophets of the credibility revolution take a dim view of functional form assumptions and instruct those who would follow them to avoid things like Millikan's bald linearization of the functional discrepancy between Stokes' law and reality for tiny drops. They also take a very dim view of selection of observations on observed values of the dependent variable, which Millikan does, and with an air of contempt for anyone who dares to disagree: '[I]t is scarcely legitimate to include such observations in the final mean', Millikan practically thunders to his reader (M11, pp. 382–383). To repeat, Millikan is all about minimizing estimation variance. If omitting those particular observations (those with speeds outside the range of ten to forty seconds) might bias his estimate, well, he is too busy trying to minimize 'likely error' to give such worries any hearing at all, whether fair or not.

I relish Millikan's clever use of both within-subject (charge) and between-subject (air pressure) treatment variation to get his estimation job done. When I read the sacred texts of the credibility revolution, it sometimes seems as if there is an unwritten but strong taboo against within-unit treatment variation. You see this most clearly when people claim the *entire* problem of social science empiricism is our inability to observe the same subject in two (or more) distinct treatments, all else held constant. I am sure this is a real problem for certain kinds of treatments, such as 'unvaccinated' and 'vaccinated'. But much fine experimental work rests solely on within-subject variation, along with a large amount of supporting work to look for violations of its own necessary conditions for validity. Millikan happily maintains the hypothesis that e_1 stays constant for any given drop as he charges, and then recharges, that drop across rounds of observation with that drop. I get the sense that a strictly constructed credibility revolution follower would not happily do so. How do I really *know* that e_1 stays constant, within drops across time, I hear them asking in arch fashion? I can only answer that I do not, but from this and similar working assumptions, a great empirical natural science was gradually built, and many Nobel laureates crowned. Empirical physics seems to get along just fine ignoring and frequently violating most of the strictures of the credibility revolution.

I also learned that Millikan's laboratory methods did not 'neutralize all forces other than the treatment' (Gerber & Green, 2012, p. 9 n. 7). It would be more correct to say that Millikan did what he could to get rid of unwanted variations in temperature, voltage, and so forth, but in the end, he carefully measured remaining variations in these crucial covariates as carefully as possible so that their remaining variance could be taken into account in constructing the apparent elementary charges e_1 . What I see in Millikan, then, is a mixed pragmatic approach: Design away whatever nuisance variance we can, and then measure and correct for what remains. It does not appear to me, then, that natural scientists have magically made some great escape from the need to take overt account of variation in disturbing causes. Perhaps the (widespread?) belief that they *have* made some such escape is a byproduct of (widespread) sloppy rhetoric

about the meaning and extent of experimental control, but I rather doubt that the scientists themselves have any illusions about this matter.

All in all, I think Millikan's empirical methods closely resemble those of conventional structural econometricians – the very people who are under assault by the credibility revolutionaries. I have no trouble couching Millikan's basic statistical problem in the homely terms of ordinary least squares. His design choices, methods, and obsessions are all very easy to understand from that banal structural econometrics perspective – whether we agree perfectly with them or not. Like a conventional structural econometrician, Millikan lets his inferences depend on pragmatic assumptions, such as a linearization of an unknown function. Like a conventional structural econometrician, Millikan depends on measurement and overt conditioning, rather than randomized assignment, to control for disturbing causes.

I also see obvious differences between Millikan's empirical practices and those of all economic empiricists, whether revolutionary or conventional. The role of sharable empirical knowledge stands out. Especially in his first experiment, Millikan depends heavily on others' measurements – specifically, for the viscosity of air and how it varies with temperature. Millikan is able to draw on 10 published studies of the viscosity of air. Complaints that economists do not have any constants are common, but notice the *conditional* nature of physicists' so-called constants: To know the viscosity of the air in his apparatus, Millikan needs not only prior estimates of an intercept and a slope, but also the temperature in his apparatus. Temperature itself had to be 'invented' across several centuries (Chang, 2007): We may not even now have the relevant covariates for doing the kind of conditioning we might like to. But I come away from Millikan with a strong sense that we would not have rich empirical social sciences until we have knowledge that travels between studies, even if only conditionally in the same way that the viscosity of air travels between physical laboratories conditional on temperature.

Although much of the rhetoric of the credibility revolution focuses on unbiased estimation, it is fair to say that most users of its tools are more interested in hypothesis-testing than measurement. To put it somewhat differently, unbiased estimation is mostly in the service of an ultimate hypothesis test. Millikan's experiments are mostly about measurement for its own sake and, in turn, Millikan depends very heavily on a large number of studies measuring the viscosity of air. I have trouble seeing how an empirical social science obsessed with piecemeal hypothesis-testing, rather than measurement and its useful generalization across contexts, will ever achieve broad empirical success and wide applicability. I have trouble seeing how an obsession with the statistical significance of causal effects brings us much closer to generalizable knowledge, if we cannot reliably condition the size of causal effects on reliably measurable covariates.

Disclosure statement

No potential conflict of interest was reported by the author.

References

- Chang, H. (2007). *Inventing temperature: Measurement and scientific progress*. Oxford: Oxford University Press.
- Franklin, A. D. (1981). Millikan's published and unpublished data on oil drops. *Historical Studies in the Physical Sciences*, 11, 185–201.
- Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. New York, NY: Norton.

- Harrison, G. A. (2013). Field experiments and methodological intolerance. *Journal of Economic Methodology*, 20, 103–117.
- Holton, G. (1978). Subelectrons, presuppositions, and the Millikan-Ehrenhaft dispute. *Historical Studies in the Physical Sciences*, 9, 166–224.
- Millikan, R. A. (1911). The isolation of an ion, a precision measurement of its charge, and the correction of Stokes' law. *The Physical Review*, XXXII, 349–397.
- Millikan, R. A. (1913). On the elementary electrical charge and the Avogadro constant. *Physical Review*, 2, 109–143.
- Smith, V. L. (1962). An experimental study of competitive market behavior. *Journal of Political Economy*, 70, 111–137.