

Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/110768/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Morey, Richard ORCID: <https://orcid.org/0000-0001-9220-3179>, Homer, Saskia ORCID: <https://orcid.org/0000-0003-1399-4895> and Proulx, Travis ORCID: <https://orcid.org/0000-0003-3763-3138> 2018. Beyond statistics: accepting the null hypothesis in mature sciences. *Advances in Methods and Practices in Psychological Science* 1 (2) , pp. 245-258. 10.1177/2515245918776023 file

Publishers page: <https://doi.org/10.1177/2515245918776023>
<<https://doi.org/10.1177/2515245918776023>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.
See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



1 Beyond statistics: accepting the null hypothesis in mature sciences

2 Richard D. Morey¹, Saskia Homer¹, & Travis Proulx¹

3 ¹ School of Psychology, Cardiff University

4 Author Note

5 This draft was compiled at Fri Apr 13 10:08:29 2018 (NA). Text, code, and data from
6 Michelson and Morley (1887) and Mendel (1866) are also available are available at
7 <https://github.com/richarddmorey/nullHistoryAMPPS>. All figures are licensed **CC BY**
8 **4.0**. Author Contributions: The main concept of the paper is due to author RM, as are the
9 two physics examples; author SH primarily contributed the genetics example. Both RM and
10 SH developed the commentary on these examples. Author TP contributed the philosophy of
11 science elements.

12 Correspondence concerning this article should be addressed to Richard D. Morey,
13 School of Psychology, 70 Park Place, Cardiff, UK. E-mail: richarddmorey@gmail.com

14

Abstract

15 Scientific theories explain phenomena using simplifying assumptions: for instance, that the
16 speed of light does not depend on the direction in which the light is moving, or that the
17 height of a pea plant depends on a small number of alleles randomly obtained from its
18 parents. The ability to support these simplifying assumptions with statistical evidence is
19 crucial to scientific progress, though it might involve “accepting” the null hypothesis. We
20 review two historical examples where statistical evidence was used to accept a simplifying
21 assumption (rejecting the luminiferous aether and genetic theory) and one where the null
22 hypothesis was not accepted in spite of repeated failures (gravitational waves), drawing
23 lessons from each. We emphasize the role of the scientific context in the acceptance of the
24 null: accepting the null is never a purely statistical affair.

25

Keywords: null hypothesis, philosophy of science, statistics

Beyond statistics: accepting the null hypothesis in mature sciences

On a warm summer morning in 1887, Albert Michelson hunched over a heavy stone table in a basement of Western Reserve College. He peered through an eyepiece whose other end disappeared under a wooden hood covering the table. With his right hand, he slowly turned a screw to calibrate one of sixteen mirrors fixed to the stone. Beneath the hood, beams of yellow sodium light bounced back and forth between the mirrors along two perpendicular paths that both ended at the eyepiece. By adjusting the screw, Michelson ensured that the lengths of the two paths were equal.

The stone slab sat on a piece of wood which itself was floating in a pool of liquid mercury. Around noon Michelson gave the table a push, causing it to slowly spin. Every 22.5 degrees of rotation — about as many seconds — he looked through the eyepiece and scribbled down a number. That afternoon he took over one-hundred readings, stopping only to give the table a small push to keep it spinning. He came back that evening for another hundred measurements, repeating the process again over the next two days.

The numbers Michelson and his colleague Edward Morley scribbled down in 1887 would eventually be among the most celebrated results in science. What they found — or rather, what they didn't find — was a quandary for popular nineteenth century theories of light propagation. Michelson and Morley's (1887) result foreshadowed not one but *two* revolutions in physics — special relativity and quantum theory — and eventually won Michelson the Nobel prize in physics.

It has been noted for decades that psychological science largely rests on the assertion of statistical differences using null hypothesis significance tests rather than resting on understanding sameness, patterns, or regularity (see e.g. Gigerenzer, Krauss, & Vitouch, 2004; Meehl, 1978; Sterling, Rosenbaum, & Weinkam, 1995). We present three historical vignettes involving null inferences (or lack thereof) in mature sciences. None of these inferences rest on significant differences from null hypothesis significance tests, but they are nevertheless examples of scientific progress. The first is Michelson and Morley's null result;

53 the second, Mendel's famous (and controversial) genetic experiments (Fisher, 1936; Mendel,
54 1866); and the third, the recent Nobel-prize-winning findings by the Laser Interferometer
55 Gravitational-Wave Observatory (LIGO) team. Understanding how the scientific context
56 supports null inferences is key to understanding why statistical nulls have traditionally been
57 ignored in psychology.

58 **Kuhnian paradigms and Normal Science**

59 In the *Structure of Scientific Revolutions*, Kuhn (1962) offers a generally descriptive
60 account of how all sciences appear to have changed over time. To the extent that these
61 changes can be construed in terms of developmental progress, they follow from motivations
62 that appear common among scientific enterprises: to generate understanding of a wide range
63 of phenomena and to provide increasingly specified guides for further scientific research.
64 Kuhn illustrates this general trajectory by identifying two stages of development.

65 The first of these stages is termed *pre-paradigm* and is marked by an absence of any
66 unifying perspectives. At this stage, theories proliferate at the pace of observed effects,
67 where these theories are little more than descriptions of a given phenomenon (e.g., when X is
68 placed over a flame, Y occurs). Because these theories offer little regarding underlying
69 mechanisms, they present no clear hypotheses beyond the replication of the original effect.
70 They are essentially tautological, reflecting little in the way of general understanding of a
71 phenomenon, and unable to produce novel predictions.

72 According to Kuhn, all scientific communities eventually acknowledge this limitation
73 and gradually make their way to a standard phase of scientific inquiry. In the *Normal*
74 *Science* phase, an underlying phenomenon is hypothesized to manifest in the various
75 previously unrelated empirical phenomena (e.g., space-time or genes). This hypothesized
76 unifying phenomenon lies at the core of a new paradigm, a broad nest of theoretical conceits
77 that shape predictions for future observations. Increasingly specifying the nature of these
78 conceits — theory articulation — guides the identification of novel effects (rather than the

79 generation of novel “theories”) and subsequent research efforts.

80 Kuhn’s depiction of Normal Science progress does not rest on a particular epistemic
81 school of thought (e.g., Popper, 1959). Rather, paradigms are understood to facilitate
82 progress by motivational means, insofar as they represent progress narratives that encourage
83 scientists to predict and accumulate paradigm-verifying effects (rather than persevere on
84 potentially falsifying anomalies, Popper, 1959 – see also the positive heuristic, Lakatos,
85 1970).

86 The motivational and verificationist realities of Normal Science have fundamental
87 implications for how the “null hypothesis” is interpreted, and whether or not it is “accepted.”
88 In a Normal Science setting, multiple explanatory paradigms offer competing accounts of
89 demonstrated effects and differing predictions for what may be observed in the future. When
90 hypotheses following from a given paradigm are not supported by the data, the null
91 hypotheses can be readily accepted, as the observed “null” effect may offer support for a
92 competing paradigm and represent an additional element of accumulated knowledge.
93 Alternatively, Normal Science may be dominated by a single, broad explanatory paradigm
94 that can account for the bulk of prior findings, and continues to make successful predictions
95 for demonstrable effects. In this setting, scientists may be extremely reluctant to accept a
96 null hypothesis that would challenge a paradigm that must be correct, insofar as it has been
97 otherwise verified in dozens (hundreds) of prior experiments, and because there are no other
98 options, meaning that the acceptance of the null could lead to a scientific crisis.

99 We discuss three examples of null effects demonstrated within the paradigmatic
100 context of Normal Science. The first and the third are from nineteenth and twenty-first
101 century physics, respectively; the second, from nineteenth- and twentieth-century biology. In
102 each case, we emphasize the relationship of the statistical inference for or against a null
103 hypothesis in the context of the relevant paradigm. Following this, we contrast the situation
104 in Normal Science with that in present-day psychology.

105 Michelson, Morley, and the luminiferous aether

106 For many centuries, there were two competing theories explaining the behavior of light.
107 Emission theory, championed by Newton, held that light was made up of particles that
108 moved in straight lines called rays. The opposing view, developed by Huygens, held that
109 light was a wave. In the eighteenth century, the emission view was dominant. Emission
110 theory is perhaps most consistent with our everyday observations of light; light appears to
111 move in straight lines, as a particle would.

112 In the beginning of the nineteenth century the wave theory of light gained the upper
113 hand among physicists due to the discovery of interference phenomena. When two waves of
114 different phases meet, they cancel and reinforce one another in complicated patterns. Light
115 behaves this way: when light is forced through slits, the light from one slit interferes with
116 light from the other, and vice versa. Interference phenomena cannot be easily explained by
117 an emission theory.

118 Expectations for light waves were built on other waves that people understood: waves
119 in water or air. If light was a wave, it must be a wave in some medium. Whatever this
120 medium is, it carries starlight above the earth and torchlight below it. It must be able to
121 pass through solid matter as light moves through glass, and it must exist in a vacuum.
122 Wave-theorists gave this mysterious medium a name: the *luminiferous aether*.

123 Physicists thought that a sea of luminiferous aether existed throughout space, providing
124 a fixed reference against which everything moves. As the earth revolves around the Sun, it is
125 passing through the aether. Facts known at the time ruled out the idea that the aether was
126 dragged along with the Earth; hence, the Earth must be moving through the aether at some
127 speed.

128 But at what speed? This was the question Michelson and Morley sought to answer.
129 Michelson had invented and refined an ingenious experimental device now known as a
130 Michelson interferometer. The 1887 version is shown in Figure 1, in both perspective view
131 (A) and top-down view (B).

132 The basic idea behind the Michelson interferometer is that it light comes from a
133 common source (Figure 1B, at a) and is focused by a lens. The light is split (b) and sent
134 along two perpendicular paths, where each beam bounces back and forth between sets of
135 mirrors. A final mirror along each path (e and e_i) sends each beam back the way it came.
136 The beams are recombined at b and pass to the eyepiece (f). The lengths of the
137 perpendicular paths can be made equal by carefully adjusting a mirror along one of the
138 paths (e_i).

139 When Michelson looked into the eyepiece while he was sending white light into the
140 interferometer, he saw a pattern of vertical dark and light bands, called “fringes”, formed by
141 the interference between the various components of white light. After calibration, Michelson
142 would rotate the stone table on which the interferometer was set. If one imagines the Earth —
143 and with it, the interferometer — moving through the aether, this rotation changes how the
144 two arms are moving with the aether “wind”. At some point in the rotation, one arm will be
145 facing into the wind, and the other arm perpendicular to it; at another point, the opposite.

146 The light moves with the aether, but the interferometer itself moves with the Earth. If
147 one arm is moving parallel to the aether wind and the other perpendicular to it, the light
148 beams in the two arms move different distances. Any difference between the arms will cause
149 the interference fringes to shift to one side by an amount that depends on the speed of the
150 Earth’s motion through the aether. Based on the 30 km/s speed of the Earth in its orbit,
151 Michelson and Morley expected the fringes to shift by a maximum of 0.4 fringe widths. This
152 maximum shift would occur when one arm is facing into the aether wind and the other
153 perpendicular to it. The minimum shift was 0, when both arms face into the aether wind at
154 the same angle (see the top of Figure 2).

155 Michelson (or Morley) gave the table a slow but steady spin and measured the shift at
156 16 rotation angles, which worked out to once every 23 seconds. They repeated the process
157 consecutively six times, at noon and in the evening, on three different days. The fringe shift
158 measurements were detrended to remove the effects of ambient temperature changes, and

159 then averaged. Michelson and Morley expected a sine curve with amplitude 0.4 fringe widths;
160 Figure 2 shows what they found.

161 There does not appear to be any discernable relationship between the angle of the
162 table's rotation and the fringe shift. There was so little effect relative to the expected 0.4
163 fringe shifts that they did not show the expected effect in their figure at all; the maximum
164 value in their figure is $1/8$ of the predicted value, because showing the predicted value in the
165 figure would hide all the variability in the data. In spite of the smallness of the effect,
166 Michelson and Morley did not directly "accept" the null. Instead, they say that

167 "[T]he displacement to be expected was 0.4 fringe. The actual displacement was
168 certainly less than the twentieth part of this, and probably less than the fortieth
169 part. But since the displacement is proportional to the square of the velocity, the
170 relative velocity of the earth and the ether is probably less than one sixth the
171 earth's orbital velocity, and certainly less than one-fourth. . . It appears, from all
172 that precedes, reasonably certain that if there be any relative motion between
173 the earth and the luminiferous ether, it must be small. . ." (Michelson & Morley,
174 1887, p. 341)

175 Indeed, this result would continue to be refined for decades using more precise
176 interferometers, and at different times of the year.¹ Michelson and Morley's result is
177 remembered as having established that there was no aether. Why is Michelson and Morley's
178 result considered convincingly null, even though Michelson and Morley merely report an
179 upper bound on the possible speed of the Earth moving through the aether?

180 **A highly-sensitive experiment.** Michelson and Morley's 1887 experiment was
181 actually the second such experiment that Michelson published. Michelson (1881) presented
182 similar results, but using a device $1/10$ as sensitive.² Other researchers noted that even

¹A recent replication by Eisele, Nevsky, and Schiller (2009) used an interferometer 100 million times as precise as Michelson and Morley's device. The result was still null.

²Michelson's 1881 paper is a model of scientific transparency. A sizeable portion of the paper is taken up

183 before accounting for a calculation mistake, “[the fringe shift] to be measured. . . was already
184 barely beyond the limits of the errors of experiment” and hence “the conclusion
185 drawn. . . might well be questioned.” Thankfully, the 10-fold increase in sensitivity was
186 possible due to a clever arrangement of mirrors. The resulting high sensitivity made for a
187 more convincing null result.

188 **A parametric manipulation.** When we discuss null results in psychology, we often
189 refer to a single effect that is not statistically significant. Michelson and Morley, however,
190 were looking for a data pattern, rather than a single effect. The sine wave pattern expected
191 due to the rotation of the table — a parameteric manipulation of the size of the expected
192 “effect” — did not present itself. The test of the theory was therefore much stronger than it
193 would have been if only one rotational angle had been considered.

194 **A theoretical expectation.** The speed of the earth moving around the sun provided
195 a value against which the null result could be compared. Michelson and Morley admit that it
196 is possible that other motion might come into play besides the Earth moving around the sun
197 — for instance, the sun moving through the galaxy — but to get such a null result, these
198 motions would all have to add up *just right* to cancel out. This would be quite the
199 coincidence, and so Michelson and Morley conclude that “chances are much against it.” They
200 note, however, that repeating the experiment at longer time intervals would allow testing
201 this possibility.

202 **Competing paradigms.** As previously mentioned, in the nineteenth century the
203 wave theory of light was dominant, but was not the only theory. The competing emission
204 theory had no need for aether. Emission theory continued to be modified to account for new
205 evidence into the late nineteenth and early twentieth century (e.g. Ritz, 1908).

describing various difficulties encountered in using his first experimental apparatus. Interestingly, although the first paper is based on results from a considerably less precise instrument, Michelson’s earlier conclusions are more definitive: “The interpretation of these results is that there is no displacement of the interference bands. The result of the hypothesis of a stationary ether is thus shown to be incorrect, and the necessary conclusion follows that the hypothesis is erroneous.”

206 Additionally, neither of the two major twentieth century theories in physics required
207 the luminiferous aether. Einstein's special theory of relativity (Einstein, 1905) made the
208 aether redundant, and quantum electrodynamics (Feynman, 1985) accounted for all the wave
209 properties of light without needing a propagation medium.

210 These four factors — the highly-sensitive experiment, the parametric manipulation of
211 the expected effect, a result far below a theoretical expectation, and a competing theory able
212 to account for the effect — combine to create the most important null result in the history of
213 science. In making the luminiferous aether unnecessary, Michelson and Morley's results
214 allowed physics move forward without it.

215 **Nuller than null: the case of Mendel and Fisher**

216 Gregor Mendel, a monk of seemingly impeccable character, conducted his famous
217 experiments on peas over the years from 1856 to 1863. The painstaking task of breeding
218 thousands of plants and carefully classifying their offspring paid off when the resulting data
219 provided evidence that genetic traits were passed on in discrete forms. Mendel's evidence was
220 close agreement of the data from his pea plants with his theory's predictions (Mendel, 1866).

221 Although Mendel's work on inheritance filled a key gap in nineteenth century biological
222 understanding, it went largely unnoticed until the turn of the twentieth century when his
223 results were rediscovered by several biologists (Piegorsch, 1990). The rediscovery sent ripples
224 through the genetics community due to its theoretical importance. A small number of
225 readers, however, noticed something else. Statistically speaking, the results were good;
226 *surprisingly* good, in fact.

227 Should a good fit to a true theory be surprising? As Pilgrim (1984) puts it, "Mendel's
228 results agreed with his theory. Why shouldn't they, since his theory was correct?" Fisher
229 (1936) took a different view. He believed the results were *too* good, and that this was
230 evidence of data falsification. Even worse, Fisher suggests that this possibly "contravene[s]
231 the weight of the evidence supplied in detail by his paper as a whole" (p. 132). This is not to

232 say Mendel was wrong, but that his results — which we review subsequently — were not as
233 evidentiary as they might initially appear.

234 Mendel's experiments considered seven traits of the garden pea plant. Pea plants, like
235 all living things, have visible traits called phenotypes that are defined by genes. For instance,
236 a pea plant's seeds might be round or wrinkled, depending on its genes. These genes come in
237 pairs — one from each parent — and can be of different forms, called alleles.

238 A dominant allele can override a recessive allele such that an organism with both types
239 of allele will have the dominant trait. The round seed shape is dominant over the wrinkled
240 shape. This means a seed with one of each allele, called heterozygous, will be round. The
241 three possible genotypes and their corresponding phenotypes are shown in Figure 3.

242 Mendel theorised there was a 50% chance of a parent passing each of its two alleles to
243 its offspring. This leads to easily predictable genotypic ratios for the seed shape of offspring
244 from two heterozygous parents (shown in Figure 4).

245 The key to Mendel's experiments were the ratio of phenotypes from crossings of
246 different plants. Mendel could infer that a plant was heterozygous if, as a seed, it was round,
247 yet some of its seeds were wrinkled. Wrinkled-seed offspring are a giveaway that the parent
248 plant must be passing on a recessive allele, and hence it *must* be heterozygous. As Figure 4
249 shows, if one crosses a heterozygous plant with itself, Mendel's theory predicts that 75% of
250 the seeds should be round.

251 Table 1 shows the Mendel's results from crossing heterozygous plants. Of 7324 seeds,
252 we would expect 5493 to be round. Mendel reports that 5474 were round, only 19 round
253 seeds from the number expected. Of course, the results of such experiments are variable: if
254 Mendel is right, the standard deviation of the number of round seeds of 7324 is
255 $\sqrt{7324 \times .75 \times .25} \approx 37$. Mendel's results are only half a standard deviation from the
256 theoretical value. By itself, this closeness is not enough to raise suspicion: there would be a
257 fair chance — 38% — of obtaining a closer result under Mendel's theory.

258 In 1936, Fisher considered all of Mendel's experiments. For every experiment, we can

Table 1

Seed totals, N , and counts of seeds with the dominant phenotype, y , for the seed shape and seed colour experiments taken from Mendel (1886); p is the theoretical proportion of seeds with the dominant phenotype predicted by Mendel's theory; z is the number of theoretical standard deviations between the expected count and observed count.

	p	N	y	Np	$y - Np$	$SD(y - Np)$	$z = (y - Np)/SD(y - Np)$
Shape	0.75	7,324	5,474	5,493.00	-19.00	37.06	-0.51
Colour	0.75	8,023	6,022	6,017.25	4.75	38.79	0.12

259 compute a deviation from the theoretical value, in standard errors. Because we are interested
 260 in the overall *distance* from the theoretical value, we square every deviation and sum them
 261 across all experiments. The result can be thought of as a squared distance, in standard
 262 errors, from the theoretical value. For round/wrinkled experiment considered above, we
 263 results were $z_1 = .51$ standard errors below the theoretical value. In a second experiment,
 264 Mendel found that 6022 of 8023 seeds contained yellow, rather than green, seed leaves. The
 265 expected proportion was 75%, or about 6017 yellow leaves. This observation is five above
 266 what was expected, a mere $z_2 = .12$ standard errors from the theoretical value.

267 We might think of the theoretical value like the bull's eye of a target, as shown in
 268 Figure 5A. The natural metric of the target is given by the expected variability of the
 269 estimate of the proportion, the standard error. The figure shows the standard errors as
 270 circles around the bull's eye. To assess how close our two experiments are to the bull's eye,
 271 we work out the distance from the center to the point $(.51, .12)$, the number of standard
 272 errors our two experiments are away from the theoretical. In the case of our two experiments,
 273 this can be found by the familiar Pythagorean theorem: $\sqrt{.28}$.

274 The distance by itself does not tell us whether the results are surprisingly close; to do

275 this, Fisher compared the observed values to the sampling distribution under Mendel's
276 theory. If Mendel was right, the squared distance for two points has a χ^2 distribution with
277 two degrees of freedom, as shown in Figure 5B. For each dimension (here, seed shape and
278 color) we expect to be somewhat off center. The more dimensions the greater the expected
279 distance, because each dimension contributes to the distance from the center. The expected
280 squared distance for two experiments is 2 (these are the degrees of freedom of the χ^2). The
281 *observed* squared distance is much smaller: .28. Our observed distance from the bull's eye is
282 closer than what we would expect 87% of the time, if Mendel's theory is correct. While far
283 from definitive, this seems close enough to cause some suspicion. But this analysis only
284 includes two of the 84 experiments reported by Mendel.

285 Fisher tabulated the results of all 84 Mendel's experiments. For clarity of presentation,
286 in Figure 6 we have grouped the related results into the 16 series suggested by Edwards
287 (1986) (Table 2, pp. 306-308), ranging from 2 to 20 degrees of freedom.³ Notice how most of
288 the squared distances from the theoretical predictions seem to be on the low side, closer to 0
289 than what we would expect. Across all 84 of Mendel's experiments, we would expect on
290 average a squared distance of 84. The observed squared distance is substantially less: 49.15.
291 To understand how small this value is, Figure 7 shows a χ^2 distribution with 84 degrees of
292 freedom, the sampling distribution of the squared distance across all experiments assuming
293 Mendel's theory. The observed distance is so small that we would expect 99.9% of such sets
294 of experiments to yield a larger distance. The experiments are *very close* to the theoretical
295 values.

296 So what? Is Weldon (1902) right when he says that Mendel's results "admirably in
297 accord with his experiment" (p. 235)? Is Pilgrim (1984) right to wonder what the fuss is all

³The two experiments we considered are series 1 in Figure 6. The results are not exactly the same as shown in Figure 5B due to the fact that Edwards (1986) has removed data that were used in another series in order to make the data in each experiment independent from the others. This also causes the overall test of all 84 experiments to be different from that computed by Fisher, but the difference does not affect the conclusions. See Edwards (1986) pp. 299-300.

298 about that results closely agree with a theory? Or is Fisher right when he suggests that
299 “most, if not all, of the experiments have been falsified so as to agree closely with Mendel’s
300 expectations” (1936, p. 132)? *Do results that agree too closely with a theoretical null actually*
301 *undermine the evidence?*

302 The last prominent statistician to weigh in on the debate was Edwards (1986), who
303 said that

304 “If it were just a question of having hit the bull’s eye with a single shot we might
305 conclude [...] that Mendel was simply lucky, but when a whole succession of
306 shots comes close to the bull’s eye we are entitled to invoke skill or some other
307 factor.” (Edwards, 1986, p. 303)

308 Of course “skill” cannot overcome the problem of inherent random variability. Both
309 Edwards⁴ and more recently Franklin (2008) suggest that Fisher’s analysis has stood the test
310 of time: Mendel’s results are too good to be true. Yet the controversy is largely unknown
311 outside of statistical circles. Why?

312 **Justified suspicion that a result is tainted does not mean it is wrong.** We
313 are in the lucky position a century and a half later of knowing that Mendel was right. Science
314 is not always neat; biases will creep into even the most rigorous research, if only because it is
315 scientific progress requires interpreting the results of experiments *post hoc* with incomplete
316 information. As (Dobzhansky, 1967) wrote at the centennial of Mendel’s publication,

317 “Few experimenters are lucky enough to have no mistakes or accidents happen in
318 any of their experiments, and it is only common sense to have such failures

⁴Interesting and relevant to the modern debate over significance testing is the fact that even the likelihoodist Edwards was persuaded by Fisher’s logic, in spite of his skepticism of significance tests. He said that “[i]t may be helpful if I admit at this point that for many years I supposed that Fisher’s analysis was going to be able to be faulted because of its total reliance on the ‘repeated sampling’ logic of the X^2 goodness-of-fit test which I had come to mistrust, but a complete review of the whole problem has now persuaded me that his ‘abominable discovery’ must stand.” (1986, p. 310)

319 discarded. The evident danger is ascribing to mistakes and expunging from the
320 record perfectly authentic experimental results which do not fit one's
321 expectations." (Dobzhansky, 1967, p. 1588)

322 Luckily Mendel described his experiments in sufficient detail that they can be easily
323 repeated. Doubt about any claim can be put to rest by rigorous replication of the procedure,
324 provided that the theory is defined clearly enough to decide what a "replication" would be.
325 Providing this clarity is one of the roles of a scientific paradigm.

326 **Interpretation of results occurs in the context of scientific theory.** This
327 seems especially obvious in the case of Mendel, given that the null was derived from
328 Mendel's theory. But suppose Mendel were a fair-minded experimentalist, and we could
329 travel back in time and confront him with Fisher's findings? Should Mendel abandon his
330 theory? Probably not. Although Fisher's critique threatens the evidential force of Mendel's
331 experiments, Fisher (1936) himself points out that Mendel, or anyone else in the nineteenth
332 century, could have derived genetic theory from three simple postulates (1936, pp. 123-124);
333 he also believed that Mendel may have done so. Fisher thought it possible that Mendel's
334 experiments were a "carefully planned demonstration of his conclusions" (Fisher, 1936, p.
335 124), rather than their sole support. Mendel's theory was strong enough to withstand
336 Fisher's critique of the evidence, in contrast to more recent psychological results subjected to
337 similar scrutiny (see e.g. Simmons & Simonsohn, 2017).

338 **Unbelievable nulls: LIGO and gravity waves**

339 Michelson's experiments using interferometers were not only important for their results;
340 the Michelson interferometer is a tool that continues to be used in research. Michelson's
341 interferometers were about 1 meter wide. Modern interferometers range from palm-sized and
342 small enough to fit in a satellite (Shepherd et al., 1993) to the immense Laser Interferometer
343 Gravitational-Wave Observatory (LIGO). The LIGO project operates two interferometers,

344 each with arms 4 km long.⁵

345 The purpose of LIGO is not to find evidence for the luminiferous aether; rather, the
346 LIGO team is hunting for gravitational waves. In Einstein’s general theory of relativity,
347 gravity is the result of changes in the geometry of space-time: a mass, such as a star, bends
348 space-time around it. When masses accelerate in certain ways — for example, black holes
349 orbiting one another — these distortions are supposed to cause gravitational waves that
350 propagate away from the source.

351 The search for gravitational waves serves two purposes: as a test of general relativity,
352 and as new way of conducting astronomy. We can use gravity waves in much the same way
353 as we use x-ray, visible-light, microwave, and radio astronomy to piece together a picture of
354 the history of the universe. Unlike light, however, gravitational waves are difficult to detect,
355 because they involve extraordinarily subtle effects as they pass.

356 This is where Michelson’s interferometer plays a key role. Laser light is split, shot
357 down the 4 km length of the two arms, bounced back from precisely suspended mirrors. The
358 laser light is recombined and passed to a detector. If the arms are the same length, the two
359 recombined waves cancel; no laser light is detected. When a gravitational wave passes an
360 interferometer, the two perpendicular arms will change lengths (Figure 8). If one arm is
361 longer than the other, then the cancelation is imperfect and some of the light makes it to the
362 detector. Space-time distortion from a passing gravitational wave shows up as fluctuations in
363 the amount of laser light at the detector.

364 Because fluctuations can happen for reasons other than gravitational waves, LIGO uses
365 multiple sites to crosscheck its results: one in Washington and one in Louisiana. LIGO also
366 cooperates with the smaller, 3 km Virgo interferometer in Italy (Figure 9). The LIGO team
367 looks for “unusual” events that occur across the detectors. Looking for correlations across

⁵Even LIGO will soon be eclipsed: the European Space Agency plans three satellites that will form an gravitational-wave-detecting interferometer with arms 2.5 *billion* meters long, called the Laser Interferometer Space Antenna (LISA). Imagine Michelson’s astonishment if he learned that the fiddly instrument with which he struggled in a Potsdam cellar would one day be built on an interplanetary scale.

368 these sites allows noisy fluctuations in only one detector to be discounted.

369 LIGO's first attempt at detecting gravitational waves in 2002 yielded a null result: that
370 is, it was deemed consistent with background noise (LIGO Scientific Collaboration, 2004).
371 Interestingly, this was expected; the first run was before the detectors were at full sensitivity.
372 The introduction to the paper is worth quoting directly:

373 "The first detection of gravitational wave bursts requires stable, well understood
374 detectors, well-tested and robust data processing procedures, and clearly defined
375 criteria for establishing confidence that no signal is of terrestrial origin. None of
376 these elements were firmly in place as we began this first LIGO science run;
377 rather, this run provided the opportunity for us to understand our detectors
378 better, exercise and hone our data processing procedures, and build confidence in
379 our ability to establish the detection of gravitational wave bursts in future
380 science runs. Therefore, the goal for this analysis is to produce an upper limit on
381 the rate for gravitational wave bursts, even if a purely statistical procedure
382 suggests the presence of a signal above background." (LIGO Scientific
383 Collaboration, 2004, pp. 102001–3)

384 Unlike Michelson's conclusion from his 1881 experiment, the LIGO team was unwilling
385 to accept the null on the basis of a noisy experiment; like Michelson and Morley's 1887
386 experiment, the LIGO state their results in terms of placing an upper limit on a quantity of
387 interest.⁶

388 From the first failure followed more. Six additional runs over more than a decade
389 would yield no evidence — at least none the team was willing to accept as inconsistent with
390 background noise — of graviational waves. LIGO became "advanced LIGO" as the team
391 improved the sensitivity of their instruments. With each failure using a more sensitive

⁶It is difficult to imagine a prominent psychology journal publishing a null result from an experiment whose purpose is to advance understanding of a methodology. Such a result would almost certainly be rejected as unimportant.

392 device, a new upper limit was established. The titles tell the story: “Upper limits on
393 gravitational-wave bursts in LIGO’s second science run” (The LIGO Scientific Collaboration,
394 2005); “Upper limits on gravitational wave emission from 78 radio pulsars” (LIGO Scientific
395 Collaboration, 2007); “Improved Upper Limits on the Stochastic Gravitational-Wave
396 Background from 2009-2010 LIGO and Virgo Data” (LIGO and Virgo Collaboration, 2014).
397 This work spawned about 100 papers from 2004 to 2016, characterizing the instruments,
398 algorithms and their improvements, or presenting data from their science runs.

399 Finally, in 2016 the team published a paper announcing the detection of gravitational
400 waves from the merger of two black holes (LIGO Scientific Collaboration and Virgo
401 Collaboration, 2016). We are more interested in what happened in the years before the
402 detection. Why were the LIGO team unwilling to accept the null and hence the possibility
403 that there were no gravitational waves? What was the difference between Michelson and
404 Morley’s situation in the late 19th century and the LIGO team’s situation in the early 21st?
405 We believe there are several.

406 **The prospect of more sensitive experiments.** The LIGO team was constantly
407 improving their instruments, and knew that more sensitive tests were just around the corner.

408 **Strong theoretical expectations and low sensitivity** The LIGO team knew
409 early on that their instruments were not sensitive enough to detect many gravitational wave
410 events of interest, should they exist. Unlike Michelson and Morley, LIGO’s null results were
411 not unexpected from the theory.

412 **No theoretical rival.** Einstein’s general theory of relativity has withstood numerous
413 tests over the past century. There is no rival to the theory that could take its place should
414 gravitational waves not exist. Plunging a field into crisis is not something to be taken lightly,
415 particularly at the expense of such a well-established theory.

416 These three conditions made the acceptance of the null hypothesis difficult, even on the
417 basis of multiple “failed” LIGO runs. Luckily, the persistence paid off. Since the 2016
418 detection, the team has made several new detections. The ability to consistently detect and

419 characterize gravitational waves has the potential to usher in a new era of gravitational wave
420 astronomy, which would not have happened if the team had accepted the null and given up.

421 **Conclusion**

422 In these three examples, a type of statistical null was rejected or accepted in relation to
423 pragmatic considerations of what would facilitate the accumulation of scientific knowledge.
424 Michelson and Morley’s result, for instance, appeared more compelling because an
425 alternative to wave theory could account for the result. On the other hand, there is no
426 alternative to general relativity, so the lack of gravitational waves would throw physics into
427 crisis. Fisher noted that Mendel could have derived his predictions from three simpler
428 theoretical postulates, rather than from the data themselves. In all three cases, the
429 evidential value of the data was considered along with higher-level theoretical concerns
430 within a theoretical paradigm. The experiments were not meant to show an isolated effect;
431 rather, they were tests or demonstrations of aspects of a broad theory.

432 In contrast, paradigmatic research programs — with concordant null hypotheses —
433 have become scarce in the contemporary field of psychology. The paradigmatic progress
434 exemplified by these three examples would not be possible within psychology’s current
435 research landscape, which closely aligns with the Kuhnian description of a Pre-paradigm
436 Science. This was not always true; in the mid twentieth century, psychological theorising had
437 coalesced into several broad paradigmatic perspectives (e.g., Cognitive Dissonance Theory,
438 Festinger, 1957). However, the subsequent decades saw psychology transform back into a
439 discipline more clearly characterized by a pre-paradigm population of micro-theories. Often,
440 these micro-theories consist solely of the described effect, followed by the word “theory” or
441 “model”, resulting in empty restatements. Insofar as they can be construed as unfalsifiable,
442 one might call them pseudo-theories (Fiedler, 2004). To the extent that these descriptive
443 theories are arrived at entirely post-hoc, they can constitute entire pseudoscience disciplines
444 (Lakatos, 1970).

445 Consider the facial feedback hypothesis (Strack, Martin, & Stepper, 1988), in which
446 feedback from the face is assumed to modulate emotion. Wagenmakers et al. (2016) recently
447 attempted to replicate the 1988 study, obtaining a null result across several labs and
448 thousands of participants. In Normal Science, this might lead to a paradigmatic crisis or new
449 boundary conditions, either of which could be construed as progress. Instead, Wagenmakers
450 et al. (2016) simply claim a failure to replicate, leaving Strack (2016) to offer a series of *post*
451 *hoc* reasons why it might not have replicated. It is not clear what was learned from the
452 episode, because the facial feedback hypothesis is not strongly linked to a broader paradigm
453 positing boundary conditions and mechanisms; it is a label for an effect. When an effect
454 stands on its own, rather than in relation to a paradigm, the implications a null result has
455 for the progress of psychological science are unclear.⁷

456 On the surface, psychology espouses the same standards of hypothesis testing as most
457 mature sciences: a Popperian (1959) emphasis on falsificationism predicated hypotheses
458 derived from explanatory paradigms. These typically constitute clear predictions that
459 distinguish the underlying explanatory accounts of distinct paradigms, allowing for the
460 specification and testing of theoretical boundary conditions which illuminate the cases in
461 which a particular paradigm may be more or less explanatory compared to its rivals
462 (McGuire, 2013). Within the context of psychology, theoretical boundary conditions
463 necessarily take on a different character, given that “theories” are often little more than
464 descriptions of phenomena, with “boundaries” that cannot extend beyond descriptions of
465 individual effects.

466 When falsification can no longer be tethered to the boundary conditions of explanatory
467 paradigms, Popperian null hypothesis testing shifts to the reliability of individual effects; if
468 an effect does not replicate as predicted, it has been “falsified” (Ferguson & Heene, 2012).

⁷This is not to say that null results are not important outside of Normal Science; it is just to say that their interpretation depends on having a paradigmatic background against which to understand them. Science is more than a catalog observations, but such a catalog may be a crucial ingredient to a developing scientific paradigm.

469 Accordingly, the historical emphasis on theory-framed hypothesis testing has been replaced
470 by the statistical significance of hypothesized effects (e.g., Benjamin et al., 2018; Open
471 Science Collaboration, 2012, 2015; Simmons, Nelson, & Simonsohn, 2011) where these
472 predictions are increasingly tested against a pre-registered hypothesis for predicted outcomes.
473 In a Normal Science setting, experimental hypotheses follow from predictions that
474 themselves follow from well-developed theories, obviating the need for the pre-registration of
475 the hypothesis. Moreover, replacing paradigmatic falsifiability with replicability of *effects*
476 discourages researchers from attending to the paradigmatic principles that allow for
477 contextualized assessments of the evidential value of a given of replication “failure” (Stroebe
478 & Strack, 2014). This further entrenches, rather than opposes, the pre-paradigm nature of
479 much of psychological science (also see Fiedler, Kutzner, & Krueger, 2012).

480 Acting within Normal Science, all three groups of experimenters we have highlighted —
481 Michelson and Morley, Mendel, and the LIGO team — are celebrated for their careful
482 experimentation. Michelson invented multiple iterations of his device to reduce the noise in
483 his measurements. Mendel grew thousands of pea plants across 84 experiments to
484 demonstrate his theory. The LIGO team invested a decade honing their experimental skills
485 before finding a single gravitational wave. This attention to detail is possible when scientific
486 progress is not defined by arguments over individual effects and statistical significance, but is
487 rather guided by work within, or opposing, a broad paradigm. If psychology reasserts itself
488 as a Normal Science, only then will it become a field unified by wide-ranging theoretical
489 perspectives in which evidence for statistical regularities are valued at least as much as
490 significant differences.

References

- 491
- 492 Benjamin, D. J., Berger, J., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., . . .
493 Johnson, V. (2018). Redefine statistical significance. *Nature Human Behaviour*, *2*, 6.
- 494 Dobzhansky, T. (1967). Looking back at Mendel's discovery. *Science*, *156*, 1588–1589.
- 495 Edwards, A. W. F. (1986). Are Mendel's results really too close? *Biological Reviews*, *61*,
496 295–312.
- 497 Einstein, A. (1905). Zur elektrodynamik bewegter körper [On the Electrodynamics of
498 Moving Bodies; M. Saha, trans.]. *Annalen Der Physik*, *17*, 891–921.
- 499 Eisele, C., Nevsky, A. Y., & Schiller, S. (2009). Laboratory test of the isotropy of light
500 propagation at the 10^{-17} level. *Phys. Rev. Lett.*, *103*, 090401. Retrieved from
501 <https://link.aps.org/doi/10.1103/PhysRevLett.103.090401>
- 502 Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias
503 and psychological science's aversion to the null. *Perspectives on Psychological Science*,
504 *7*(6), 555–561. Retrieved from <https://doi.org/10.1177/1745691612459059>
- 505 Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.
- 506 Feynman, R. P. (1985). *QED: The strange theory of light and matter*. Princeton University
507 Press. Retrieved from <http://www.jstor.org/stable/j.ctt2jc8td>
- 508 Fiedler, K. (2004). Tools, toys, truisms, and theories: Some thoughts on the creative cycle of
509 theory formation. *Personality and Social Psychology Review*, *8*(2), 123–131.
510 Retrieved from https://doi.org/10.1207/s15327957pspr0802_5
- 511 Fiedler, K., Kutzner, F., & Krueger, J. I. (2012). The long way from α -error control to
512 validity proper: Problems with a short-sighted false-positive debate. *Perspectives on*
513 *Psychological Science*, *7*(6), 661–669. Retrieved from
514 <https://doi.org/10.1177/1745691612462587>
- 515 Fisher, R. A. (1936). Has Mendel's work been rediscovered? *Annals of Science*, *1*(2),
516 115–137. Retrieved from <http://dx.doi.org/10.1080/00033793600200111>
- 517 Franklin, A. (2008). The mendel-fisher controversy: An overview. In A. Franklin, A. W. F.

- 518 Edwards, D. J. Fairbanks, D. L. Hartl, & T. Seidenfeld (Eds.), *Ending the*
519 *mendel-fisher controversy* (pp. 1–77). Pittsburg, PA: University of Pittsburg Press.
- 520 Gigerenzer, G., Krauss, S., & Vitouch, O. (2004). The null ritual: What you always wanted
521 to know about significance testing but were afraid to ask. In D. Kaplan (Ed.), *The*
522 *Sage handbook of quantitative methodology for the social sciences*. Thousand Oaks,
523 CA: Sage.
- 524 Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago
525 Press.
- 526 Lakatos, I. (1970). History of science and its rational reconstructions. In *Proceedings of the*
527 *biennial meeting of the philosophy of science association* (pp. 91–136).
- 528 LIGO and Virgo Collaboration. (2014). Improved upper limits on the stochastic
529 gravitational-wave background from 2009-2010 LIGO and virgo data. *Phys. Rev.*
530 *Lett.*, *113*, 231101. Retrieved from
531 <https://link.aps.org/doi/10.1103/PhysRevLett.113.231101>
- 532 LIGO Scientific Collaboration. (2004). First upper limits from LIGO on gravitational wave
533 bursts. *Phys. Rev. D*, *69*, 102001. Retrieved from
534 <https://link.aps.org/doi/10.1103/PhysRevD.69.102001>
- 535 LIGO Scientific Collaboration. (2007). Upper limits on gravitational wave emission from 78
536 radio pulsars. *Phys. Rev. D*, *76*, 042001. Retrieved from
537 <https://link.aps.org/doi/10.1103/PhysRevD.76.042001>
- 538 LIGO Scientific Collaboration and Virgo Collaboration. (2016). Observation of gravitational
539 waves from a binary black hole merger. *Phys. Rev. Lett.*, *116*, 061102. Retrieved
540 from <https://link.aps.org/doi/10.1103/PhysRevLett.116.061102>
- 541 McGuire, W. J. (2013). An additional future for psychological science. *Perspectives on*
542 *Psychological Science*, *8*(4), 414–423. Retrieved from
543 <https://doi.org/10.1177/1745691613491270>
- 544 Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the

- 545 slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46,
546 806–834. Retrieved from
547 <http://www.psych.umn.edu/faculty/meehlp/113TheoreticalRisks.pdf>
- 548 Mendel, G. (1866). Versuche über Pflanzen-Hybriden (Bateson, Trans.). *Verhandlungen des*
549 *naturforschenden Vereines in Brünn*, 42, 3–47.
- 550 Michelson, A. A. (1881). The Relative Motion of the Earth and the Luminiferous Ether.
551 *American Journal of Science*, 22, 120–129.
- 552 Michelson, A. A., & Morley, E. W. (1887). On the Relative Motion of the Earth and the
553 Luminiferous Ether. *American Journal of Science*, 34, 333–345.
- 554 Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate
555 the reproducibility of psychological science. *Perspectives on Psychological Science*, 7,
556 652–655.
- 557 Open Science Collaboration. (2015). Estimating the reproducibility of psychological science.
558 *Science*, 349(6521), 943. Retrieved from [dx.doi.org/10.1126/science.aac4716](https://doi.org/10.1126/science.aac4716)
- 559 Piegorsch, W. W. (1990). Fisher’s contributions to genetics and heredity, with special
560 emphasis on the Gregor Mendel controversy. *Biometrics*, 46(4), 915–924. Retrieved
561 from <http://www.jstor.org/stable/2532437>
- 562 Pilgrim, I. (1984). The too-good-to-be-true paradox and Gregor Mendel. *Journal of*
563 *Heredity*, 75(6), 501–502. Retrieved from
564 <https://academic.oup.com/jhered/article/75/6/501/876950>
- 565 Popper, K. (1959). *Logic of scientific discovery* (2002 eLibrary.). London: Routledge
566 Classics.
- 567 Ritz, W. (1908). Recherches critiques sur l’électrodynamique générale [Critical researches on
568 general electrodynamics; J. Lucier, trans.]. *Annales de Chimie et de Physique*, 13,
569 145–275. Retrieved from <http://www.datasync.com/~rsf1/crit/1908a.htm>
- 570 Shepherd, G. G., Thuillier, G., Gault, W. A., Solheim, B. H., Hersom, C., Alunni, J. M., ...
571 Harvie. (1993). WINDII, the wind imaging interferometer on the Upper Atmosphere

- 572 Research Satellite. *Journal of Geophysical Research: Atmospheres*, 98(D6),
573 10725–10750. Retrieved from
574 <http://onlinelibrary.wiley.com/doi/10.1029/93JD00227/abstract>
- 575 Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology:
576 Undisclosed flexibility in data collection and analysis allows presenting anything as
577 significant. *Psychological Science*, 22, 1359–1366. Retrieved from
578 <https://doi.org/10.1177/0956797611417632>
- 579 Simmons, J. P., & Simonsohn, U. (2017). Power Posing: P-Curving the Evidence.
580 *Psychological Science*, 28(5), 687–693. Retrieved from
581 <http://dx.doi.org/10.1177/0956797616658563>
- 582 Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited:
583 The effect of the outcome of statistical tests on the decision to publish and vice versa.
584 *The American Statistician*, 49(1), 108–112. Retrieved from
585 <https://www.tandfonline.com/doi/abs/10.1080/00031305.1995.10476125>
- 586 Strack, F. (2016). Reflection on the smiling registered replication report. *Perspectives on*
587 *Psychological Science*, 11(6), 929–930. Retrieved from
588 <https://doi.org/10.1177/1745691616674460>
- 589 Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the
590 human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of*
591 *Personality and Social Psychology*, 54(5), 768–777.
- 592 Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication.
593 *Perspectives on Psychological Science*, 9(1), 59–71. Retrieved from
594 <https://doi.org/10.1177/1745691613514450>
- 595 The LIGO Scientific Collaboration. (2005). Upper limits on gravitational wave bursts in
596 LIGO's second science run. *Phys. Rev. D*, 72, 062001. Retrieved from
597 <https://link.aps.org/doi/10.1103/PhysRevD.72.062001>
- 598 Wagenmakers, E.-J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., R. B. Adams, J., ...

- 599 Zwaan, R. A. (2016). Registered replication report: Strack, martin, & stepper (1988).
600 *Perspectives on Psychological Science*, 11(6), 917–928. Retrieved from
601 <https://doi.org/10.1177/1745691616674458>
- 602 Weldon, W. F. R. (1902). Mendel's laws of alternative inheritance in peas. *Biometrika*, 1,
603 228–254.

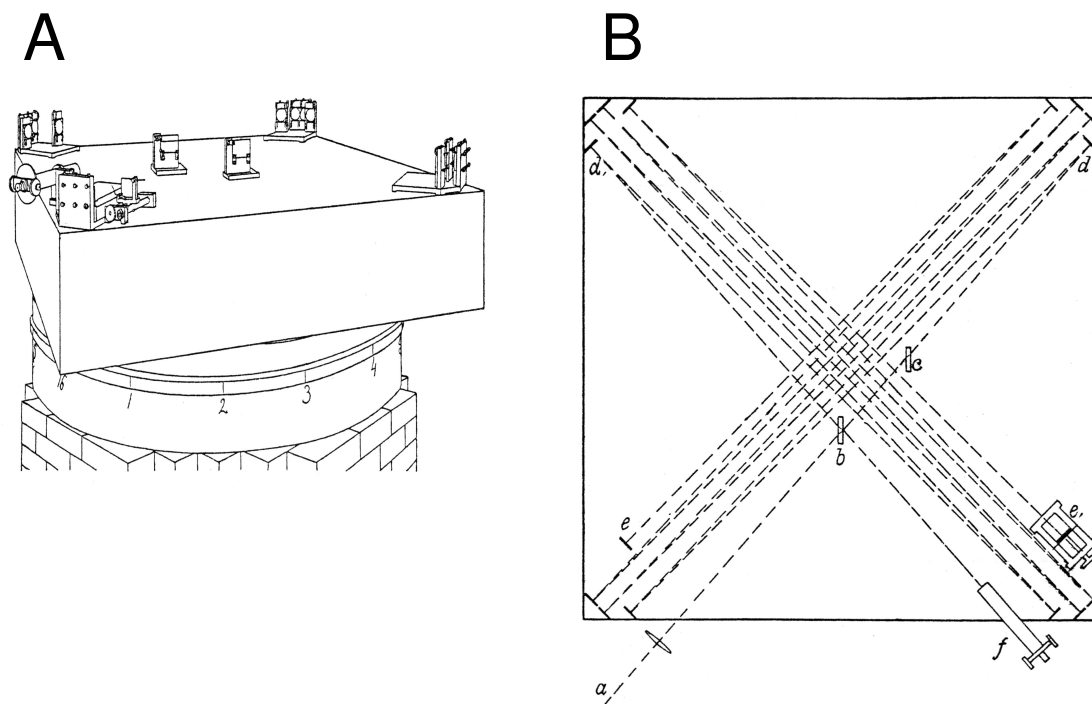


Figure 1. Michelson and Morley's device (1887, fig. 3 and 4 from the manuscript). A: Perspective drawing of the device without its wooden cover. The surface was about 1.5m square. B: Schematic of the table surface. Light emitted from the light source a through a lens hits a beam splitter b and is sent along one of two perpendicular paths. The light is then reflected back and forth by mirrors at d and d_i (and opposite), until they are reflected back by mirror e or e_i . They pass back through the beam splitter and part of both beams is sent to an eyepiece at f . The mirror e_i is finely adjustable so that the two beams can be equated in length. An extra beam splitter c is used to ensure that both beams move through the same amount of glass.

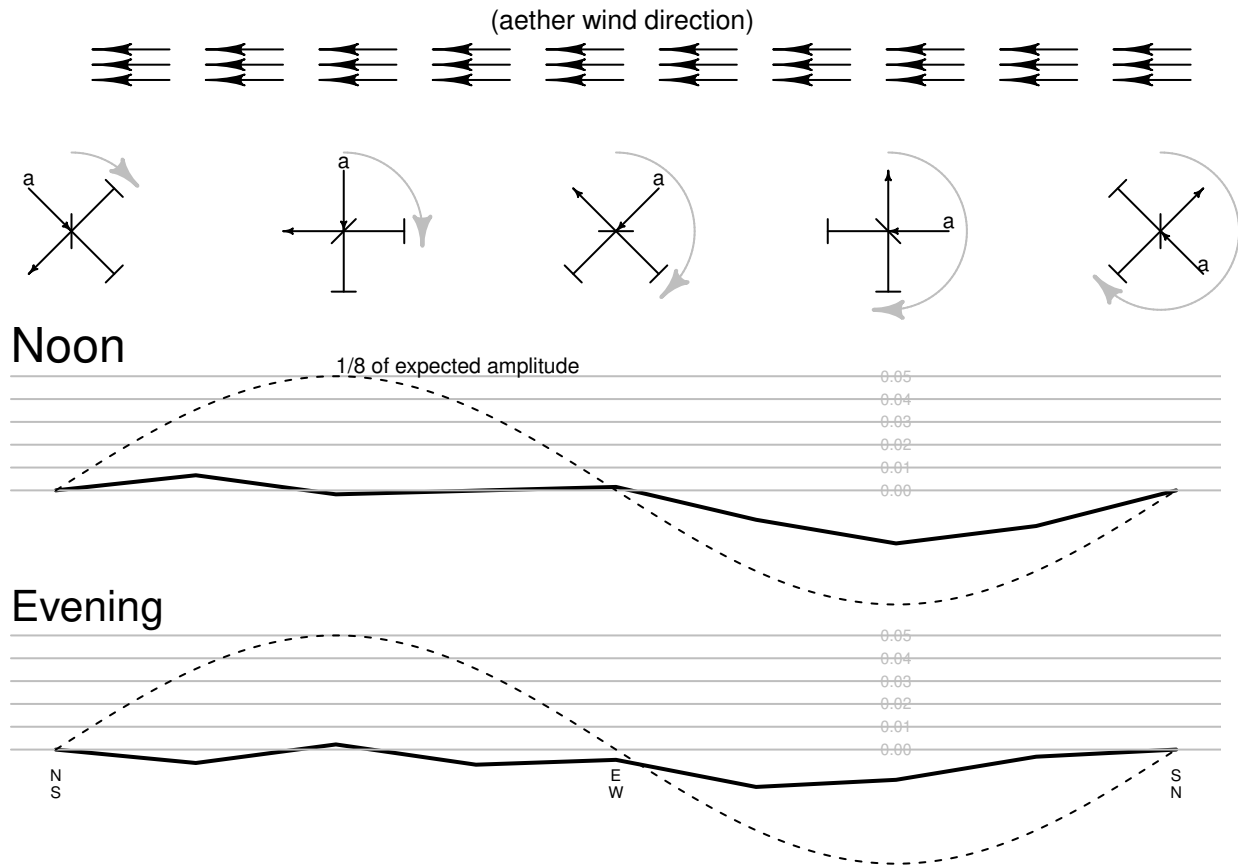


Figure 2. The data from Michelson and Morley’s experiment, as presented in the manuscript. The top series shows the average of the detrended noon runs, and the bottom the detrended evening runs. The y axis is the amount of shift in fringes. The dotted curve shows the expected pattern at $1/8$ the expected amplitude of 0.4. In the schematic above, the point marked “a” represents the light source on the sketch of the instrument.







Genotype	Phenotype	
Heterozygous		 Round
Homozygous Dominant		 Round
Homozygous Recessive		 Wrinkled

Figure 3. All possible genotypes and corresponding phenotypes for the seed shape trait. Seed shape has two possible alleles (round and wrinkled) and the round allele is dominant. Icons for the genotypes (black-and-white) and phenotypes (solid black) are shown here and used in subsequent figures. The circle denotes the round allele and the star, the wrinkled allele.

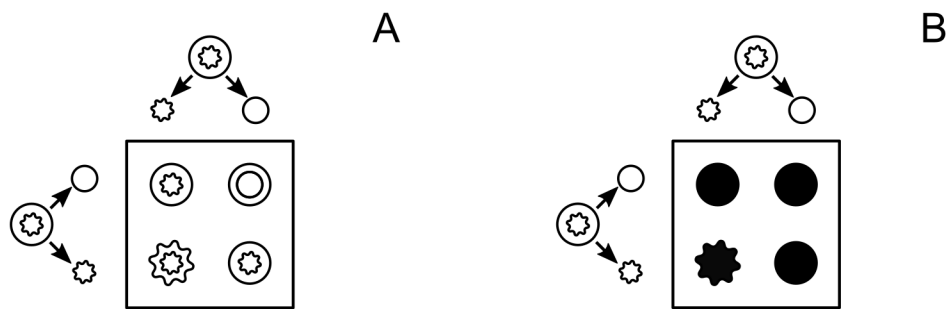


Figure 4. An example of Mendelian genetics with two heterozygous parents (left and top of each square). Inside the squares are the four crossings of the two alleles from each parent. A: The genotype of each possible cross. B: The phenotype of each possible cross. Although 50% of the alleles correspond to the wrinkled phenotype, only 25% of the resulting plants will be wrinkled due to the wrinkled allele’s recessiveness.

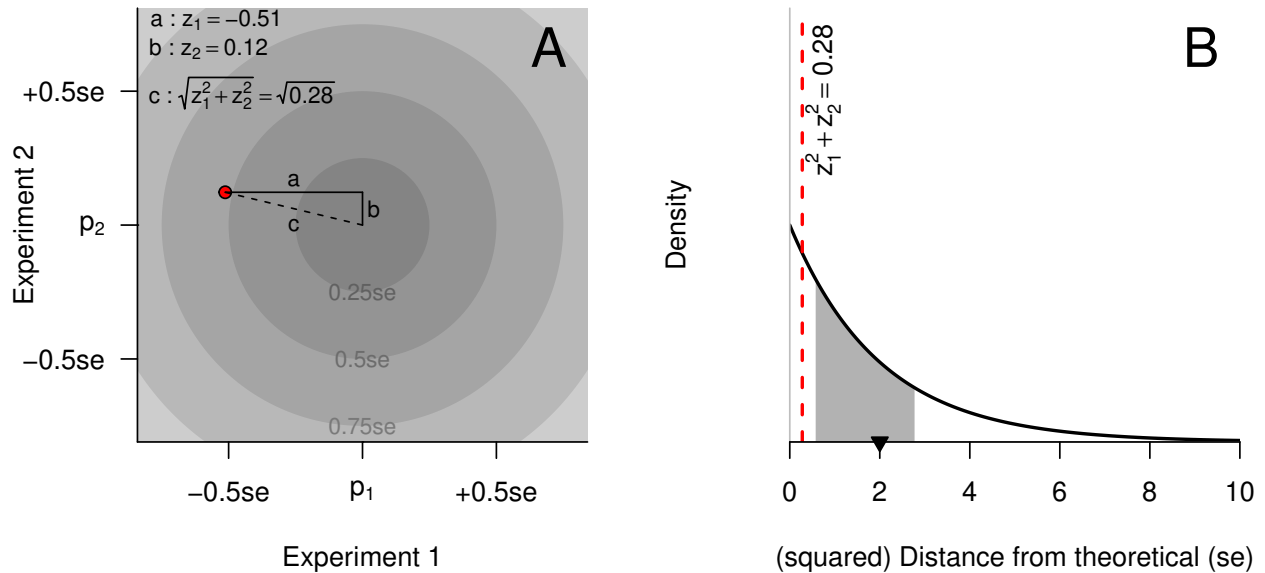


Figure 5. A: Calculating the distance, in standard errors, of a pair of estimates (red circle) from the theoretical values (center of the bull’s eye). Diamonds on the axes show the individual observations in each experiment. B: The distribution of the squared distance, assuming two points. The expected squared distance is 2, as shown by the triangle on the bottom axis. The probability of getting a smaller squared distance than the one observed is about .13, assuming Mendel’s theory. The shaded region shows the middle 50% of the distribution.

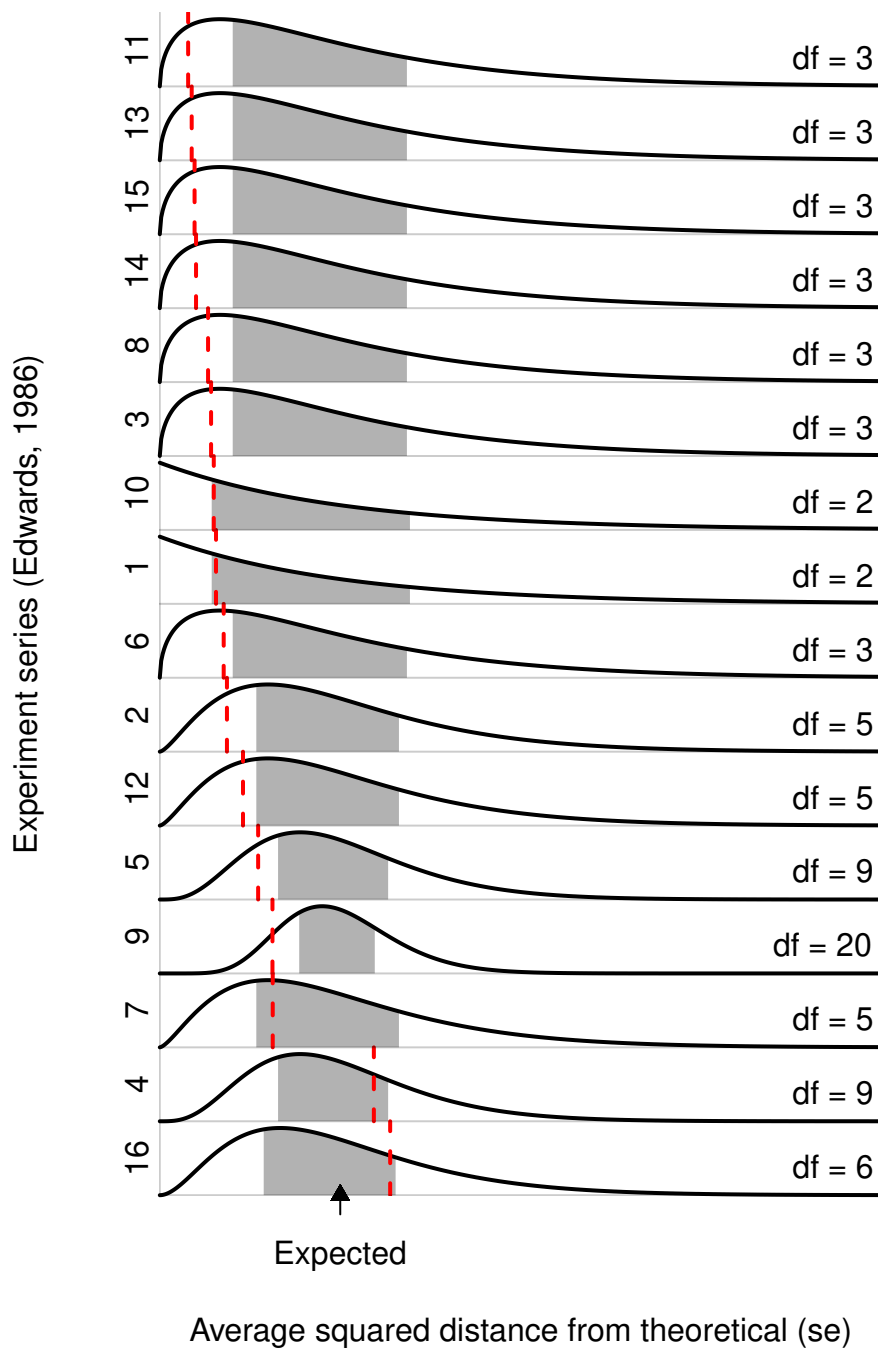


Figure 6. Results from Edwards’ (1986) sixteen groupings of Mendel’s 84 experiments, along with theoretical distributions. The series are sorted by deviation from expectation, and scaled by expectation (degrees of freedom) in order to visually align all the results. Shaded regions show the middle 50% of the distributions.

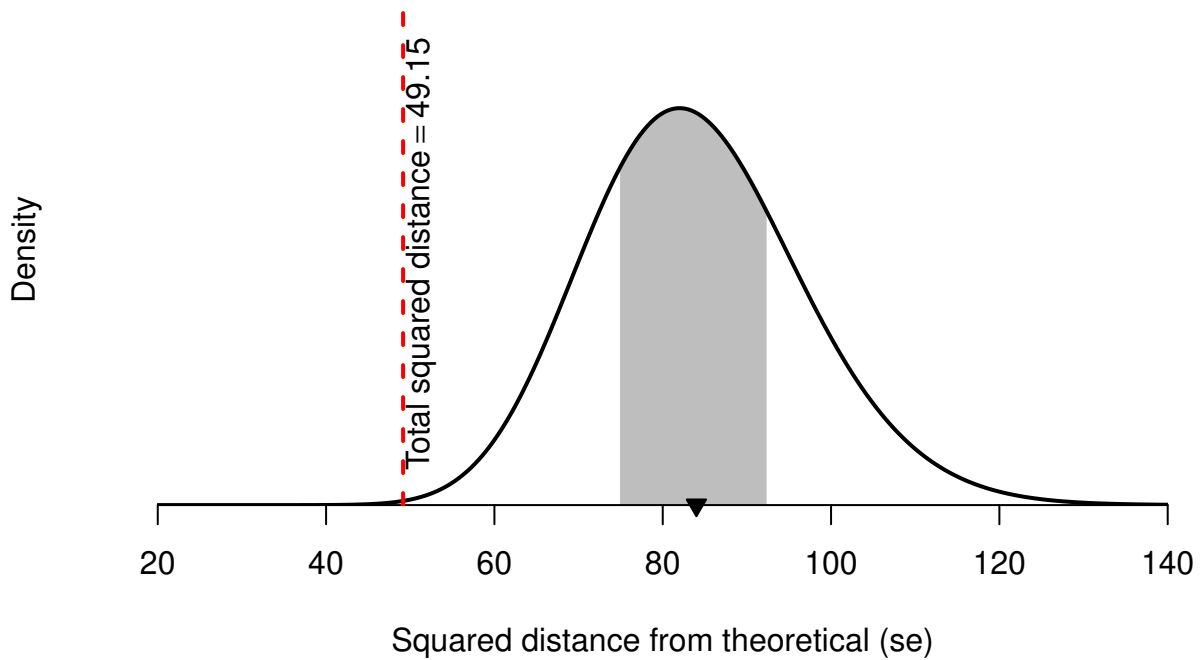


Figure 7. Theoretical distribution across all 84 experiments. The red line indicates the observed total squared distance 49.15 (calculated from Edwards’ 1986 data). There is a 99.9% chance that a random value from this distribution would be larger than 49.15. The shaded region shows the middle 50% of the distribution and its expectation is indicated by the triangle on the bottom axis.

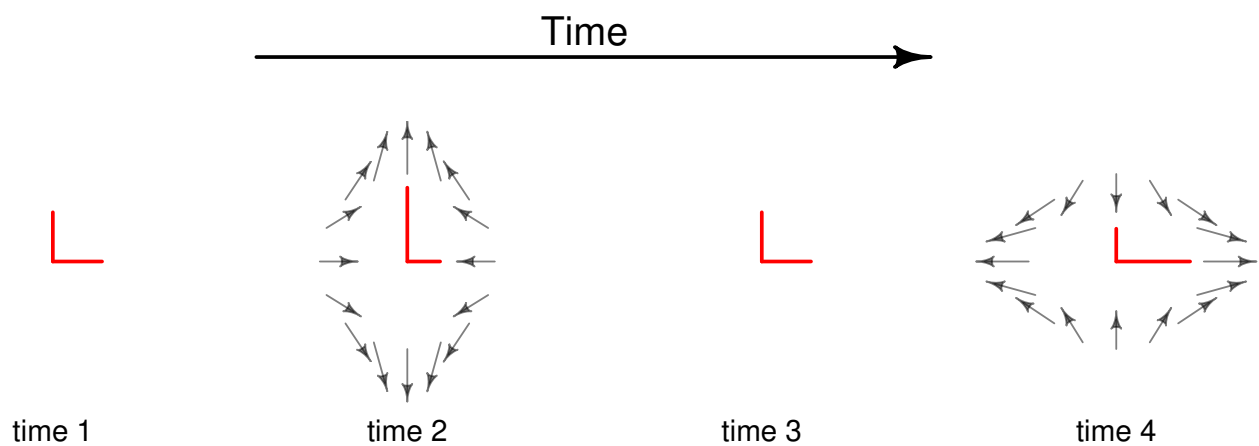


Figure 8. How gravitational waves distort the length of the two perpendicular arms of the LIGO Michelson interferometers.

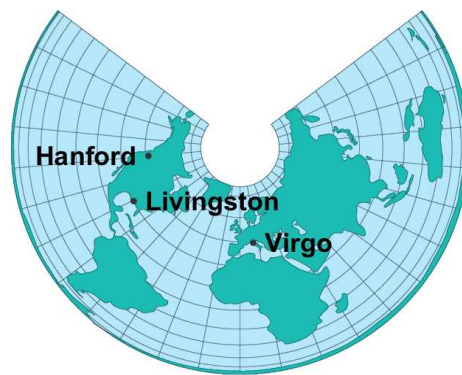


Figure 9. Locations of the Laser Interferometer Gravitational-Wave Observatory (LIGO) sites in the United States, and the Virgo interferometer in Italy.