Chapter 11: Summary and conclusions.

11.1 In search of evidence.

Not all data is gathered with the intention of identifying evidence about hypotheses (for instance), but much data is and any form of analysis that cannot extract evidence from data is of limited use. Although some of the statements made by Neyman and Pearson suggest that their techniques were not intended to identify evidence, others[1] suggest the opposite, and it is clear that their methodology is widely believed to be of value in this respect, nor is this belief confined to those with limited statistical knowledge.

Fisher's approach was intended to answer evidential questions and he criticised Neyman and Pearson on the grounds that their methodology could not encompass this aim. Two principles – each originating from the theories of Fisher – have come to be associated with the search for evidence within a frequentist framework. The first is the *sufficiency principle* and the second, some form of *conditional principle*. Both have traditionally been interpreted with respect to conventionally large parameter spaces. Each, in its way, attempts to remove from analytical influence elements of the total design that seem to be irrelevant to the evidential question at issue, and both have been incorporated into some versions of frequentist theory over the course of time.

11.2 Problems with using conventional unconditional frequentist inference to find evidence.

In order to be evidentially applicable, a statistical technique should, at the least, be able to give a sensible answer to the question: *What data is strong evidence against hypothesis H relative to hypothesis K?* Yet, even when H and K are simple

[1] Pearson (1955) quoted in Lehmann (1993), p. 1244.

hypotheses, this question creates difficulties for conventional frequentist inference. The usual answer – *data with a small p-value* – is unsatisfactory on a number of grounds, including the following.

Problem (i):
Data that is clearly much more consistent with H than K may have a small p-value.

Problem (ii):
The same data may be interpretable (according to the p-value criterion) as *both* ‘strong evidence against H relative to K’ and ‘strong evidence against K relative to H’.[2]

Problem (iii):
Data with a small p-value may be nearly equally consistent with H and K (this is particularly noticeable when the null and alternative distributions are very similar).

Problem (iv):
The p-value is very insensitive to the exact formulation of K – an unsatisfactory feature for any measure of relative evidence.

11.3 Improving frequentist inference by conditioning.

The conditional methods of Fisher and Cox (who allied it with the work of Neyman and Pearson) seem to suggest a way in which frequentist inferences can be improved, since removing parts of the sample space that have been identified as irrelevant to the question at issue may improve the quality of the inference. However, these methods, although they can produce results that are dramatically different from the unconditional, do not bring us any closer to avoiding the problems described above, nor do they mitigate the scale of these problems to even the slightest degree.

[2] Problems i. and ii. are closely related to each other but describing them in this way makes the evidential unreasonableness more apparent.

Evidently, a small conventional p-value is not a sufficient reason to reject H (in favour of K) since it does not guarantee a small likelihood ratio, but this is also true of the conditional p-values produced by Cox or Fisher's methods. For example, in Cox's two-stage example using two Normal populations, the conditional test is simply the conventional (i.e. unconditional) test that would have been applied had there been only the one (observed) population; it is therefore subject to the full force of the problems cited above.

In contrast, the more robust conditional principle, used by Birnbaum, combines with the SP to give the LP; this is satisfied by techniques[3] (such as that of Royall) that are completely free of these problems, however these techniques are not frequentist.

There is a huge gap between the effect of Cox's conditionality principle and Birnbaum's. We were intrigued by this disparity and wondered whether there was any way in which conditioning could be used, within the confines of frequentism and without breaching the SP, to remove or mitigate some of the more outstanding problems. The fact that many statisticians are deeply wedded to the frequentist approach was a major consideration is pursuing this question.

It can be argued that, in an evidential context, we ought not to test composite hypotheses since the different elements of the composite (i.e. component simple hypotheses) are likely to return varying measures of relative evidence for any given data, and it follows that no measure of evidence is valid over the whole composite. This motivated us to consider testing simple hypotheses only, that is, using binary parameter spaces. Exhaustive conditional inference (ECI) uses Cox's conditional principle (which we have called 'restricted'), exactly, but applies it to statistics that are ancillary on binary parameter spaces instead of the traditional large parameter spaces; this makes a critical difference because we have been able to identify optimal versions of such ancillary statistics for a wide range of cases. Even though this approach is unequivocally frequentist, problems (i) and (ii) never occur, and the extent of problems (iii) and (iv) is generally reduced, and, in some cases, completely overcome.

[3] Most Bayesian inferences also satisfy the LP but it is not necessary to use priors in order to achieve this.

In addition to the direct argument (above) for using simple hypotheses, it seems that using binary parameter spaces is a more effective way of extending conditioning, within the frequentist framework, than allowing the use of statistics that are *approximately* ancillary on the large parameter space. We have access to an exhaustive ancillary statistic[4], which defines a strong form of optimality, and the method is more straightforward because we do not need to specify *how close* to ancillary a statistic must be before we condition upon it – a question that raises many other questions. The 'which ancillary statistic' dilemma is exacerbated by allowing approximate ancillaries to stand in competition with non-optimal exact ancillaries, whereas the fact that ECI is optimal solves[5] this problem.

11.4 Properties of exhaustive conditional inference.

ECI mitigates or solves problems (i)-(iv).

ECI solves problems (i) and (ii) in all cases, as follows. Since the critical likelihood ratio of any test defined by a small conditional significance level is less than *one*, the rejection region for $y = \frac{f_H(\underset{\sim}{x})}{f_K(\underset{\sim}{x})}$ is always of the form $(0, r_1]$ where $r_1 < 1$. Hence we cannot observe any data that is both in the rejection region and much more consistent with H than K.[6] Since we reject K (as null) in favour of H only when $y \in [r_2, \infty)$ where $r_2 > 1$, it follows that no data can result in both the rejection of H in favour of K and the rejection of K in favour of H.

Exhaustive conditional inferences can suffer from problem (iii), for instance, in the Exponential case with $n = 1$ and $q = 1.01$. In that case the distributions of Y under H and K are very similar for $y < 1$, and data with a non-significant likelihood ratio of $\frac{1}{2}$ has a cp-value of 1%, indicating significance. This is unsatisfactory but an improvement on the conventional result, since the p-value is even lower. On the other

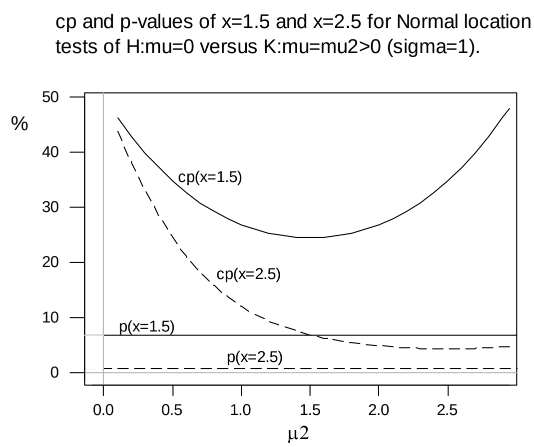
---

[4] This is restricted to cases where Y is a continuous variable and a few *ad hoc* extra cases.
[5] At least, up to the point where we have two competing exhaustive ancillary statistics, and we have yet to identify such a case when Y is continuous.
[6] Thus ECI renders redundant Kendall and Stuart's ad hoc modification to conventional inference, which was designed to solve problem (i). See Kendall & Stuart, pp. 182, 183.

hand, in the cases $q=\frac{1}{2}$ and $q=\frac{2}{3}$, where no data has a small likelihood ratio, ECI produces cp-values that are not significant, for any data, unlike the p-values. Thus, when $q=1.01$, problem (iii) is mitigated by ECI and when $q=\frac{1}{2}$ or $\frac{2}{3}$, it is solved. There is some evidence that, in any particular case, using a sufficiently large sample will solve this problem; this is certainly true in the Exponential case and may be true more generally. If so, it contrasts with conventional inference, where increasing n ultimately exacerbates the problem (see **Figure 10.18**).

With respect to problem (iv), we note that ECI is usually very sensitive to the exact formulation of the alternative hypothesis, K. Consider right-sided Normal location tests of the form H: $\mu=\mu_1$ versus K: $\mu=\mu_2$ where μ_2 can be any value greater than μ_1. The cp-value of any fixed observation, x, varies with μ_2 (and thus the cut-off value for any α-level test also varies μ_2), whereas the conventional p-value of x is constant over all $\mu_2>\mu_1$. The following diagram shows the cp-values and p-values of the observations $x=1.5$ and $x=2.5$ for testing $\mu_1=0$ versus $0<\mu_2<3$ where $\sigma=1$.

Figure 11.1



We have derived some general results about the pairing functions of DDF statistics and from them can deduce some of the general properties of ECI. However, there is much that we do not know about pairing functions and our knowledge of the general properties of ECI is similarly limited. Over and above the properties discussed above, we are left with conjectures and some general impressions of ECI based on our observations of the examples given in Chapter 10. Since these particular examples may be influential, we should say something about how they were selected.

The models we considered are those most commonly discussed and used for conventional inference. Our choice of hypotheses was based on simplicity and, where possible, we looked for cases where an explicit form could be calculated for the pairing function, as in the Exponential model with $q = \frac{1}{2}, \frac{2}{3}, \frac{3}{2}$ and 2. We spent some time looking for cases where the ECI breaks down (and found them in some scenarios with hypotheses extremely close together) – this was the only type of result that we sought out. When it came to choosing data, we always included observations where the conventional inference gives poor results, in order to ascertain whether or not ECI would improve matters. The Gradient model was developed out of a desire to find a model that would reproduce some of the useful features of Welch's Uniform model (simplicity, intuitiveness, and the ability to illustrate issues to do with universal ancillarity), with an explicit pairing function (giving the general form of the ECI), but without the complication of a discrete likelihood ratio statistic leading to breaches of the sufficiency principle.

Other desirable properties of ECI.

The DDF statistics on which ECI is based are *exhaustive*, meaning that the subset of the sample space that remains after conditioning is as small (in terms of the number of distinct y-values) as an ancillary set can ever be. It follows from this that no more conditioning can be done without the loss of ancillarity. Let D be the DDF statistic and A any other statistic ancillary on the binary parameter space. When we condition on both these statistics, we are conditioning on $A \times D$, then either $A \times D$ is equivalent to D alone, or $A \times D$ is not ancillary.

It follows from this that the DDF statistic is always a *maximal* ancillary statistic.

ECI satisfies the *sufficiency principle* for the given binary parameter space since the DDF statistic is a function of the likelihood ratio statistic, which is minimal sufficient.

(Recall that conditioning on a statistic that is ancillary on a large parameter space often breaches the SP with respect to a given BPS and thereby introduces an arbitrary element into the inference.)

A *small Type I error probability*, of the traditional kind, is frequently guaranteed by an ECI test. This is always the case when the resulting critical likelihood ratio is small, by Robbins' result (see §7.3: *Robbins' result*), and often even when it is not. Thus the Exponential scenario with $n = 1$ and $q = 1.01$ and conditional significance level of approximately 1% does not have a small CLR since $CLR \approx \frac{1}{2}$, nevertheless the unconditional significance level is almost *zero*. A small Type I error probability is a desirable before-experiment design characteristic of a test although we do not believe that it should be the basis of an after-experiment evidential inference. In none of the cases that we have examined has the ECI produced a large Type I error probability or even a Type I probability that is greater than the (appropriately small) conditional significance level, however we have not proved that this can never happen.

Continuity.

A point of *inferential discontinuity* occurs when two scenarios (including data) are arbitrarily similar but result in inferences that are not arbitrarily close. Optimal unconditional inferences are continuous because, as one scenario (or series of scenarios) tends to a limiting case, so do all the components of the inference – such as the p-value – and thus the inferences converge. However, when we condition on exact ancillary statistics, a situation can arise where a particular statistic is exactly ancillary in the limit but not elsewhere (or *vice versa*) so that the limiting inference is conditional and substantially different from the limit of the other, unconditional, inferences. Allowing an inference to be made conditional upon a statistic that is *approximately ancillary* (weakly dependent on θ) can seem like a solution to this problem, but it only moves the point of discontinuity, it does not remove it. Savage (1970) pointed out that the requirement that an ancillary statistic be a function of the MSS contributes to the problem yet ignoring this requirement causes breaches of the sufficiency principle or takes one out of frequentism. Conventional conditional inference causes many such discontinuities.

If we use conditional inference in some cases, but not in others (because no ancillary statistic exists[7]), such discontinuities will inevitably occur. How should we create a general method that incorporates ECI? We could use ECI whenever the DDF statistic in ancillary (essentially when Y is continuous and in some other *ad hoc* cases) and use either unconditional or a combination of unconditional and conventional conditional methods otherwise. In this case our overall approach will contain many discontinuities, although there will be none within the class of cases to which ECI is applicable.

Within the ECI class of scenarios, there is continuity. This follows from the fact that the ancillary statistics on which ECI is based all have the same basic structure – they are *the difference of the distribution functions of the likelihood ratio statistic*; thus, whenever a series of such scenarios converges to a limiting case also of this form, their DDF statistics must converge to the limiting DDF statistic; since their DDF statistics are all exactly ancillary, they are all conditioned upon with the result that the inferences converge. Traditional conditioning has been based on the use of ancillary statistics that have no particular structure in common (other than ancillarity) and are fairly rare; there is, therefore, no continuity claim that can be made about the class of scenarios subject to this type of conditioning.

Since we have shown that unconditional and conventional conditional methods do not (generally) produce realistic evidential measures, we might prefer to limit ourselves to using ECI where this is possible and leave other cases unaddressed. This approach creates a coherent whole but at the cost of leaving many standard scenarios outside the scope of the method. (If the sample is very large, it may be the case that the distribution of Y is very well approximated by a continuous distribution – even though actually discrete – and we might choose to use ECI as an approximation.

[7] There are many cases where no ancillary statistic exists. To create an example of such a case, simply define a variable on a very limited discrete support (e.g. $\{x_{1,} x_{2,} x_3, x_4\}$) and define two distinct and valid probability distributions on the support such that no x-value, or combination of x-values, short of the entire set, has the same total probability under both distributions.

However, care must be taken in identifying appropriate approximating distributions; traditional approximations, such as the Normal approximation to the Binomial, have been used because they approximate *tail-areas* well, rather than *densities*, and they approximate the likelihood ratio $(y = \frac{P_H(x)}{P_K(x)})$ only poorly.)

There is a third possibility, namely that we could use likelihood methods when the DDF is not ancillary. However, I take it that anyone prepared to do this is not very attached to frequentism and might as well use likelihood methods across the board.

Closer to a 'likelihood' interpretation.

Whether or not one regards closer agreement with *likelihood methods* as a point in favour of ECI depends on one's attitude to the former. However, it is interesting to note that a greater level of congruence between frequentism and the likelihood method of Royall, for instance, can be achieved by conditioning exhaustively. ECI is not consistent with the likelihood principle since it does not produce a universal relationship between the LR and the cp-value (there is more than one inference class), yet it produces inference classes of substantially greater size than those produced by other frequentist methods and the universal upper bound (*one*) on the value of the critical likelihood ratio (for any $\alpha < 100\%$) ensures that there is a limit to how much the cp-value and likelihood ratio can conflict. In many cases, most obviously those in the log-symmetric category, ECI provides a frequentist justification for a test procedure that is perfectly consistent with a likelihood outlook.

Different stopping rules will usually produce different p-values for the same data, even when the likelihood ratio of the data is the same. If we adhere to the LP, this ensures that data produced by two different stopping rules, but with the same likelihood ratio, is interpreted the same way. ECI only has this effect if both stopping rules produce the same pairing function, as sometimes happens. However, even when this is not the case, if both stopping rules produce cp-functions $(cp_1$ and $cp_2)$ that are increasing in y, then the general results $cp(y) < y$ and $\lim_{y \to 1} cp(y) = \frac{1}{2}$ ensure that

$$|cp_1(y) - cp_2(y)| \leq \min(y, \tfrac{1}{2}), \ \forall y \in (0,1).$$

While, if both stopping rules produce cp-functions that are decreasing in y, then y must be bounded below by some value $\frac{1}{2} < y' < 1$ and

$$|cp_1(y) - cp_2(y)| \leq y' - \tfrac{1}{2}, \ \forall y \in [y',1).$$

In all cases, $|cp_1(y) - cp_2(y)| = 0, \ \forall y > 1$, since any cp-value of a likelihood ratio greater than *one* is 100%.

It is also the case that $p(y) < y$, and this places a non-trivial bound on the difference between two conventional p-values when $y < 1$, however, the fact that a value of $y > 1$ may produce a small p-value means that there is no non-trivial upper bound on the absolute difference between the p-values produced by two different stopping rules when $y > 1$.

Practicality.

Although tedious to perform on an *ad hoc* basis, exhaustive conditional inference, including test and interval results, could easily be incorporated into statistical software packages. In addition, it is particularly simple to perform complete ECI – covering cp-value, significance level, power and interval estimates – on the log-symmetric class of scenarios where all these measures are simple functions of the likelihood ratio of the data, and the functions are the same for every scenario within the class (in contrast to unconditional inference, even in the Normal case). Were it to be shown that the log-symmetric class covers a great number of scenarios *asymptotically*, then (subject to some knowledge of convergence rates) this analytic simplicity would extend to a wide range of cases.

11.5 Conjectures and unanswered questions.

Asymptotic convergence of ECI.

There is a possibility that the cp-values for a wide range of scenarios may converge as $n \to \infty$, creating a large asymptotic E. C. inference class. Such a class can only include the Normal location scenario if the limiting inference is that which applies to the log-symmetric cases, i.e. $cp(y) = \frac{y}{(1+y)}$ $(y < 1)$ and in Chapter 10 we showed that this was at least a possibility. The implications of this would be considerable since it would mean that, when n is large, ECI – which is a purely frequentist method – gives something very close to a likelihood interpretation of the data. This follows from the fact that the consistent relationship between the likelihood ratio (y) and the cp-value is such that values of the likelihood ratio usually regarded as significant evidence against H relative to K correspond to cp-values that are similarly interpreted (see §8.7: *The cp-value and the likelihood ratio*).

Is the DDF statistic uniquely exhaustive?

We have not shown that the DDF statistic is the only exhaustive ancillary statistic that can be defined for these scenarios. We have limited this discussion to cases where the likelihood ratio statistic, Y, is continuous, because this is a sufficient condition for identifying an exhaustive ancillary statistic – the DDF statistic. In cases where Y is a *discrete* variable, there will usually be no exhaustive ancillary statistic, but it is possible to find examples where there are two non-equivalent exhaustive ancillary statistics. We have not identified any exhaustive statistics, other than the DDF statistic, for continuous Y, but this does not prove that such cases do not exist. Any exhaustive ancillary statistic can be defined by a pairing function; the pairing function of the DDF statistic is decreasing in y when $y<1$. *If* we could show that, for continuous Y, *all* exhaustive ancillary statistics have pairing functions that are decreasing in y when $y<1$, then it would follow, from an inversion of the proof of ancillarity in Chapter 9 (§9.2: *Proof*), that the DDF statistic is *uniquely* exhaustive. The log-symmetric models (Chapter 8) and the Gradient model (Chapter 10) have symmetry properties that allowed us to identify exhaustive ancillary statistics directly ($|\ln Y|$ and $|X|$ respectively); each of these ancillary statistics is 'equivalent' to the corresponding DDF statistic.

Is it ever the case that the cp-value is significantly small but the p-value is not?

We have come across no instances of cases where the cp-value indicates significant evidence against H, relative to K, but the conventional p-value does not. This certainly cannot happen in any of the log-symmetric cases where we have shown that the p-value is always less than the cp-value. (And, this may be true more generally if the asymptotic conjecture is true.) In both the Gradient and Exponential models, there are some cases where the cp-value is less than the p-value but, in all those cases, both values are greater than 50% so significance is not an issue. *If* it could be shown that the function $cp(y)$ on $(0,1)$ is always *either*, (a) monotonic increasing, (b) monotonic decreasing, or (c) constant[8], this would be sufficient to show that there are no cases where the cp-value, but not the p-value, is significant, as follows.

Recall that $cp(y) \to 50\%$ as $y \to 1$. Thus, for (b), the result is trivial since $(\text{b}) \Rightarrow cp(y) \geq 50\% \; (\forall y < 1)$ and hence $p(y) > cp(y) \Rightarrow \; p(y) > 50\%$ and not significant. Under assumptions (a) and (c), it is possible to show that $p(y) \leq cp(y)$ by using the same approach as the proof of this result in the log-symmetric case (§8.11) and the fact that, for all DDF statistics and $y_0 < 1$,

$$y_0(a) = \begin{cases} D_1^{-1}(a), \; a \leq D(y_0) \\ \text{non-existent, elsewhere.} \end{cases}$$

11.6 Two fundamental criticisms of ECI.

ECI allows one to condition to the maximum degree possible while still remaining frequentist. Since any conditional principle is closely connected to the likelihood principle – logically, if not psychologically[9] – it follows that conditional inference is partly motivated by elements associated with that principle, as well as with traditional frequentism. The most fundamental criticisms that can be made of ECI are those that are implicit in the pure form of either of the two intellectual traditions underlying conditional inference.

[8] We have no counter-examples to this claim.
[9] Those who are attached to a conditional principle do not necessarily have any sympathy for the LP, to which they may even be deeply opposed, yet all versions of the CP are entailed by the LP and seem to share the same 'relevance' motivation.

In the section on continuity, we noted that depending on how we incorporate ECI into frequentism, we either produce a method with many points of inferential discontinuity or a method that is less widely applicable than alternative methods. Each of the following approaches, at odds with ECI, solves both these problems.

Unconditional inference.

In the earlier chapters of this work, we were at pains to explain and defend the position sometimes called ‘the two machines argument’, which maintains that a conditional inference is superior for assessing evidence or showing ‘what the data tells us’ about a particular question. At the heart of this argument is the perceived irrelevance of outcomes within *unobserved ancillary subsets* of the sample space, which can be equated to *machines that were not used in the experiment* or *unperformed sub-experiments*. Despite what we regard as the compelling nature of this argument, it is not universally accepted. Welch, for example, dismissed the Fisherian version of the argument on the basis that the overall (average) power of a test (rather than the conditional power) was the ‘real’ power, and there is no question that the overall power of a conditional test procedure is lower than that of the optimal unconditional test with the same overall significance level. If, on this or any other grounds, one does not accept the desirability of conditioning, even when the aim of the exercise is to find evidence, then the ECI project has no appeal. This is a fundamental criticism since acceptance of the two machines argument is a primary motivation for pursuing any conditional approach. To support unconditional inference, it is necessary to justify overlooking the anomalies that it produces (such as (i)-(iv) in §11.2), either on the basis that they are not really anomalies (by disputing our implicit interpretation of ‘evidence’, for instance), or on the basis that they are a price worth paying for the advantages of unconditional inference.

ECI does not go far enough.

In response to the two machines argument, we have identified certain ancillary statistics with a view to conditioning upon them. In order to retain both the sufficiency principle and the general frequentist framework, we have defined ‘ancillary’ in a way that is more restricted than Birnbaum’s version. With this restriction in place, exhaustive conditional inference is as far as we can go towards homing in on the most relevant part of the sample space, but does it go far enough? According to the likelihood principle the only relevant element of the sample space is the data or likelihood ratio actually observed (y_0) and no other observation should influence the analysis.[10] On this view, the entire frequentist project is completely wrong-headed.

Fisher's proposed remedy [to the problem created by different data containing different levels of reliability] was not to question the orthodox reasoning which caused this anomaly, but to invent still another *ad hockery* to patch it up: use sampling distributions conditional on some 'ancillary' statistic $z(x_1,\ldots,x_n)$ that gives some information about the data configuration that is not contained in the estimator[11].

Since the two machines argument justifies the *unrestricted* conditional principle, one can argue that it should move us to abandon frequentism in favour of the likelihood principle. We then satisfy both the unrestricted CP and the SP, and an approach, such as that of Royall, is free of all the problems that we have highlighted ((i)-(iv)), simplifies the methodology by the direct use of likelihood ratios, and is inferentially continuous and generally applicable.

11.7 Likelihood-like frequentism.

This work evolved from frequentist traditions and, therefore, rules out the use of prior probabilities on hypotheses. Given this constraint, there are three major options available: (i) a reasonably pure version of Neyman-Pearson methodology, (ii) a purely Likelihoodist method, or (iii) some kind of foot-in-each-camp approach. Any approach of the third kind is likely to be motivated by a preference for conditional inferences over unconditional, and this, in turn, implies an interest in evidence. ECI

[10] In ECI, $\pi(y_0)$ influences the result.

[11] Jaynes (2003), p. 253.

satisfies these requirements best (within a frequentist framework), since it involves stronger conditioning than any other frequentist approach and, although it requires the use of binary parameter spaces, there are good independent reasons for doing this if you want to evaluate evidence. What light does the development of ECI cast on the pure theories?

On the one hand, ECI tends to confirm the view that the conventional error probabilities (and p-value) of a test do not measure anything useful, at least, from the point of view of someone primarily interested in evidence. The conventional significance level of any test of simple hypotheses is revealed as the (before experiment) *expectation* of the significance level, over an infinite number of notional sub-experiments, of varying quality, *only one of which is actually performed to produce the data.*[12]

On the other hand, ECI provides a frequentist justification for the type of results produced by non-frequentist methods such as that of Royall. The development of ECI shows that an inference method does not necessarily give results that are substantially different from likelihood results *simply by virtue of being frequentist*. We have seen many cases where ECI interprets the data in a way that is either very similar to or (allowing for the different measures used) equivalent to the interpretation based directly on the likelihood ratio value, while both are substantially different from the unconditional reading of the data, and this seems to be the case (on both counts) whenever the sample size is sufficiently large. It is thus no longer simply a matter of choosing between frequentist and likelihood interpretations, even as a first step; a much greater incongruity lies between the results obtained by conventional frequentism (including the limited forms of conditioning available to date) and exhaustive frequentism than between the latter and likelihoodism.

[12] Where the choice of sub-experiment is independent of hypothesis.