STUDIA UBB. PHILOSOPHIA, Vol. 65 (2020), 2, pp. 51-66
(RECOMMENDED CITATION)
DOI:10.24193/subbphil.2020.2.03

KNOWLEDGE-HOW, ABILITY, AND COUNTERFACTUAL SUCCESS. A STATISTICAL INTERPRETATION[1]

ADRIAN LUDUȘAN*

ABSTRACT. The paper is thematically divided into two parts. In the first part, we will address the arguments raised against the anti-intellectualist thesis that ability is a necessary condition for knowledge-how, present Katherine Hawley's proposed generic solution based on counterfactual success in order to overcome these arguments, followed by an analysis of Bengson & Moffett's counterargument to Hawley's counterfactual success thesis [CST]. We will conclude that Bengson & Moffett's counterargument misses its target, so that, as far as we are concerned, Katherine Hawley's proposal, namely CST, is safe. In the second part of the paper, we will provide a statistical interpretation of one of Hawley's more specific proposals, counterfactual success with occasional failure [CSTF], and assess a couple of philosophically challenging consequences that follow from such an interpretation.

Keywords: *know-how, ability, counterfactual success, intellectualism, anti-intellectualism, null hypothesis significance testing, effect size.*

Introduction

The epistemological discussions surrounding knowledge-how are canonically[2] divided as debates between two broad views about what knowledge-how consists

[1] I would like to thank Ciprian Bogdan, Ion Copoeru, Mihai Rusu, and Roberta Șerban for their careful reading, suggestions and comments on earlier versions of this paper. Also, I would like to express my gratitude to the organizers and participants in the Wednesday Research Seminar hosted by the Department of Theoretical Philosophy (University of Bucharest) & CELFIS, where an earlier version of the paper was presented; I greatly benefited from their helpful comments and suggestions.
* Faculty of European Studies, Babes-Bolyai University, Strada Emmanuel de Martonne 1, Cluj-Napoca 400090. Email: adiludusan@gmail.com
[2] See (Fantl, 2017)

in and what its nature[3] is. The first view posits that to have knowledge-how is, in effect, to have propositional knowledge under certain modes of presentation or ways of thinking, while the second view is characterized by the idea that to have knowledge-how is to possess a certain 'power'[4], namely a certain ability or disposition, to act accordingly. Beginning with Gilbert Ryle[5], the first view has come to be known as 'intellectualism', and the second view, as 'anti-intellectualism'.

Anti-intellectualism came in two flavors, according to the hypothesized nature of the 'power' in question: an ability, or a disposition. In the remainder of the paper we will focus on the ability account of power. Ability, in anti-intellectualism, is related to knowledge-how either as a necessary condition or as a sufficient condition. As a consequence, many intellectualists have tried to undermine one or both of these logical relationships. In order to accurately describe the positions under scrutiny here, the proposed arguments and counter-arguments, we will adopt Bengson & Moffett's[6] terminology, definitions, and specifications. First, we will consider the arguments raised against the anti-intellectualist thesis that ability is a necessary condition for knowledge-how, present Katherine Hawley's proposed solution to overcome these arguments, and the intellectualist response to Hawley's proposal. Then, we will argue that the response misses its target, so that, as far as we are concerned, Katherine Hawley's proposal is safe. In the second part of the paper we will provide a statistical interpretation of one of Hawley's proposals and assess a couple of philosophically challenging consequences that follow from such an interpretation.

Counterfactual success and knowledge how

Following Bengson & Moffett[7] we will specify the thesis that ability is a necessary condition for knowledge-how as:

[AIN] Having the ability to ϕ, or having had the ability to ϕ at some time in the past, is necessary for knowing how to ϕ.

The intellectualists arguments against AIN can be summarized in a paradigmatic example:

---

[3] The distinction between the nature and grounding of knowledge how can be tracked in (Bengson & Moffett, 2012, pp. 162-163)

[4] (Bengson & Moffett, 2012)

[5] (Ryle, 1949)

[6] (Bengson & Moffett, 2012)

[7] (Bengson & Moffett, 2012)

Ski Instructor. Pat has been a ski instructor for twenty years, teaching people how to do complex ski stunts. He is in high demand as an instructor, since he is considered to be the best at what he does. Although an accomplished skier, he has never been able to do the stunts himself. Nonetheless, over the years he has taught many people how to do them well. In fact, a number of his students have won medals in international competitions and competed in the Olympic games[8].

Now, what the intellectualists argue using the ski instructor example is that Pat knows how to do complex stunts, although he lacks the ability to do them, so, voilà, a case of knowledge-how without the corresponding ability. We think that the argument can be resisted, although we are not going to analyze the argument, explore and develop the strategies by which it can be countered. Instead, we will focus on the proposal advanced by Katherine Hawley[9]to handle such arguments. The idea underlining Katherine Hawley's proposal is that the tension between what counts as evaluable worlds with respect to ability vs knowledge-how is superficial: we tend to evaluate ability in worlds very close to the actual world, and knowledge-how at worlds which are very close to worlds similar with the actual world, but an elementary examination of the tasks involved in such cases should dissolve the apparent tension. So, a careful reformulation of AINin terms of counterfactual situations should fix the problems that the ski instructor example generates.

Counterfactual Success Thesis [CST]: x knows how to ϕ only if:
if x tried to ϕ under normal conditions, x would succeed at ϕ-ing.

To see how CST may be used to counter the argument based on the ski instructor example, one should note that once we remove the 'under normal circumstances' clause (for there is nothing abnormal in Pat's conditions), the evaluable worlds for Pat's know-how respect CST.

Bengson & Moffett concede that CST might accommodate the argument based on the ski instructor, but they think that the anti-intellectualist construal of AINin terms of CST remains problematic. To this end, they have proposed an argument based on the following situation:

Pi. Louis, a competent mathematician, knows how to find the n^{th} numeral, for any numeral n, in the decimal expansion of π. He knows the algorithm and knows how to apply it in a given case. However, because of principled computational limitations,

[8] (Bengson & Moffett, 2012, p. 168)
[9] (Hawley, 2003)

Louis (like all ordinary human beings) is unable to find the 10^{46} numeral in the decimal expansion of π[10].

The argument goes like this: 'Notice that conditions would have to be extremely abnormal for Louis to succeed in finding the 10^{46} numeral in the decimal expansion of π when he tries: he would have to be superhuman, as it were. Presumably, then, we need to consider very "distant" or "dissimilar" worlds to locate one in which Louis succeeds in his attempt. In this world, and presumably all others even remotely like it, Louis cannot reasonably hope to succeed in finding the 10^{46} numeral in the decimal expansion of π when he tries. His inability is pervasive. Yet he still knows how to find it. So the counterfactual success thesis—with or without the 'under normal conditions' clause—is false. Call this the problem of pervasive inability for the anti-intellectualist thesis that an ability to act is necessary for knowing how to act.'[11]

Under scrutiny[12], however, *Pi* turns out to be a poor argument against AIN or CST, for it misconstrues the way in which ability or counterfactual success are employed in these theses. Along the lines of Hawley's[13] analysis, one should distinguish the relevancy of the tasks in question with respect to the envisaged know-how. At close examination, we discern in *Pi* two distinct tasks, one which consists in finding the n^{th} numeral in the decimal expansion of π, the other consisting in finding the 10^{46} numeral in the decimal expansion of π; let us assume that the former task is understood in what we consider to be the usual sense[14], that

---

[10] (Bengson & Moffett, 2012, p. 170)

[11] (Bengson & Moffett, 2012, p. 171)

[12] The analysis of *Pi* that we are going to present differs from that of (Cath, 2015); although we are sympathetic to the idea underling Cath's argument against *Pi*, that 'rather than simply knowing how to Φ one always knows how to Φ -in-circumstances-C1 or Φ -in-circumstances-C2' (Cath, 2015, p. fn 14), so that the corresponding tasks for Louis are: find-the-10^{46}-numeral-in-the-decimal-expansion-of-π-in-circumstances-where-he-has-much-greater-computational-powers, and find-the-10^{46}-numeral-in-the-decimal-expansion-of-π-in-circumstances-where-he-has-his-current-computational-powers, we do not agree with Cath's verdict that Louis knows how and has the corresponding ability with respect to the first task, but neither with respect to the second: we fail to understand why Louis doesn't know-how to find-the-10^{46}-numeral-in-the-decimal-expansion-of-π-in-circumstances-where-he-has-his-current-computational-powers, given that he has mastered the algorithm.

[13] See Hawley's analysis of the difference between assessments of knowledge-how vs ability in (Hawley, 2003, p. 23).

[14] In order to show that this is the usual sense, consider the case of multiplication for the same situation: Louis knows how to multiply two natural numbers, and we think there is consensus that he has the corresponding ability (being 'a competent mathematician'), but, obviously, he is unable to find the result of the multiplication of two natural numbers, each composed of 10^{46} digits. Are such extreme cases powerful enough to make us reconsider the attribution of ability to Louis? Our answer is that we will still maintain that Louis has the generic ability of multiplying two natural

is, for reasonable values[15] of n. The first task, then, has a generic character, and assuming the possession of the corresponding know-how, is achievable under normal human conditions and circumstances, while the second task, all things being equal, is not. From this point on, we will qualify as 'relevant' a task that, assuming the possession of the corresponding know-how, is achievable under normal human conditions and circumstances, and as 'irrelevant' one whose achievement, assuming the full possession of the corresponding know-how, depends on conditions and circumstances that go beyond what is humanly achievable. The two above-mentioned tasks determine two corresponding abilities or counterfactual successes: the generic ability or counterfactual success of/in finding the n^{th} numeral in the decimal expansion of π, and the specific ability or counterfactual success of/in finding the 10^{46} numeral in the decimal expansion of π; obviously, of these two pairs of abilities and counterfactual successes, only the first one qualifies as relevant. We should emphasize that only relevant tasks, abilities, and counterfactual successes are the subject matter of CST and AIN, under a proper construal of these theses.

In light of this distinction, one can observe that Louis's inability to find the 10^{46} numeral in the decimal expansion of π is based on 'principled computational limitations', and, so, is irrelevant to his know-how. The relevant ability in Louis's case is the ability to find the n^{th} numeral in the decimal expansion of π, and we can safely assume from the characterization of Louis as 'a competent mathematician' that he possesses it: we assume that he can successfully determine the n^{th} numeral in the decimal expansion of π, given a reasonable value of n with respect to his computational capacities. Of course, his know-how is not limited by his computational capacities, after all he mastered the algorithm, but the fact that his corresponding relevant ability is, has, again, nothing to do with the ability of determining the decimal expansion of π: if Louis's computational capacity and memory were to miraculously be extended, with no extra requirements besides those that he already possesses – know how + corresponding relevant ability – Louis would succeed in determining the 10^{46} numeral.

In order to emphasize our last point, let us return to the ski instructor example, and suppose that Pat knows not only how to do complex ski stunts, but also has the ability to do them. Now, although it is uncontroversial that Pat has the ability to do the complex ski stunts (we have assumed this), Pat is unable to perform the 10^{46th} stunt. However, as we can observe, his inability has nothing to do with

numbers, even when confronted with such extreme cases, because we implicitly understand that such an ability involves only reasonable values of natural numbers with respect to human computational capacities.

[15] Reasonable with respect to human computational capacities, that is.

his know-how or the generic ability to do complex ski stunts; it is a pseudo-inability, so to say, based on inherent human limitations: nobody has the ability to do something which requires drastically transcending normal human conditions, although they have the ability to perform the same type of actions within the area circumscribed by their human limitations. In CST terms, counterfactual success should be evaluated only if it is humanly achievable, that is, under normal human conditions, there is at least one accessible nearby possible world or counterfactual situation where success is achievable.

There is another point worth mentioning which reinforces our verdict: ability, we reckon, as employed in AIN, has a generic character, which explains its versatile applicability in different contexts, so we regard as unwarranted any numerical requirement associated with an ability, such as, for example, the ability to shoot the 10^{46th}target, the ability to play the 10^{46th}chess game, or the ability to find the 10^{46} numeral in the decimal expansion of π.

Consequently, *Pi* should be judged by the same standards as the *Ski instructor* example: Louis knows how to find the decimal expansion of π and has the corresponding generic ability within the limits defined by his cognitive conditions, just as we agreed that Pat knows how to do complex ski stunts, has the corresponding generic ability, but, obviously, and unrelated to his knowledge-how or generic ability, he is unable to perform the 10^{46th} stunt.

Technically, our analysis amounts to considering that counterfactual success or the possession of the relevant ability with respect to knowledge-how should be attainable under normal human conditions, that is, the accessibility relation that circumscribes the range of possible worlds with respect to the relevant ability or counterfactual success should be defined by human attainability. Success in counterfactual worlds, or ability, is relevant for know-how only if there is at least one proximal world at which success is attainable. As a consequence, 'pervasive inability', that is, inability or lack of success in all accessible counterfactual worlds is irrelevant and, thus, should be disregarded, having no bearing whatsoever on AIN or CST. Obviously, Louis's inability to find the 10^{46} numeral in the decimal expansion of π occurs at all worlds accessible from its human vantage point, so its corresponding 'inability' bears nothing against CST. In conclusion, we think that the argument against CST based on *Pi* misses its target by misconstruing 'ability' and 'counterfactual success' in AIN and CST. In conclusion, CST should be amended by the stipulation that ϕ should be relevant or accessible.

So far, we have presented our defense of Hawley's account of knowledge-how against the so-called 'problem of pervasive inability', next we will focus on developing a positive reinforcement of her proposal by appeal to certain well-known and widely used statistical techniques.

Counterfactual success with occasional failure. A statistical interpretation

In her paper 'Success and Knowledge-How', Hawley presents two objections against her strategy of rendering AIN in terms of CST. The first one is a variant of the ski instructor, which we have already shown how to dismantle, and the second is based on the occurrence of failures even in the presence of certified or at least agreed knowledge-how. Hawley proposes and shortly discusses two options[16] for handling such cases: the first is to define what a subject knows how to do in terms of just those circumstances in which the subjects succeeds, so that under those circumstances, success is guaranteed, and the second option is to allow some room for failure, arguing that what counts is that the subject would usually succeed in performing the task.

We are going to focus on the second option described by Hawley, but not before explaining our reasons for endorsing this option. And the reasons have to do with the shortcomings of the first option. The first reason concerns the circular characterization of knowledge-how under this option: if x knows how to ϕ under circumstances c, then, if x tried to ϕ, under circumstances c, x would successfully ϕ, where c = circumstances in which x succeeds in ϕ-ing. Of course, as Hawley argues[17], the circularity involved here is benign, for two reasons: first, knowledge-how and counterfactual success must be supplemented with justification[18] in order to link them[19], and, second, often, competence is defined, without a pernicious logical fallacy, with respect to competent performers: if x is a competent performer, then x succeeds under c, where c = circumstances in which a competent performer succeeds. Still, that leaves us with an unrealistically fine-grained conception of

[16] Hawley doesn't explicitly endorse one of the two proposed solutions, although when comparing them, she describes the epistemological virtue of the first option as more desirable: 'The first option has the advantages of theoretical simplicity, retaining a straight counterfactual success condition, whilst allowing some leeway in which task is discussed. The second option introduces some fuzziness into the standards for knowledge how, rather than into what is known" (Hawley, 2003, p. 24).

[17] (Hawley, 2003, p. 24)

[18] Hawley's analysis of knowledge-how mirrors the traditional epistemological analysis of knowledge-that; she puts true belief and knowledge in correspondence with counterfactual success and knowledge-how, so that the link between true belief and knowledge, namely justification, is mirrored by a link between counterfactual success and knowledge-how; she refers to this latter link as 'warrant'. However, she sometimes writes as if knowledge-how must be supplemented with warrant in order to explicate counterfactual success, for example, when she says that '[s]he may, if she also satisfies some "warrant" condition, know how to …' (Hawley, 2003, p. 24) or 'knowledge-how requires warrant as well as success' (Hawley, 2003, p. 24).

[19] So, to characterize one in terms that involves itself is not 'viciously circular' as she (Hawley, 2003, p. 24) puts it.

knowledge-how: 'She may, if she also satisfies some "warrant" condition, know how to make delicious bread when ingredients are available, she has normal use of her body, she is concentrating, the oven doesn't break down half way through baking, ... there is no need to fill in the dots.'[20]

The second reason is that under this option failure is impossible, which, we consider, is unrealistic, and diminishes the heuristic character of mistakes and failures in the process of learning, even in the case of experts. We are not going to supplement our remarks with studies, empirical confirmations, and arguments, suffice to note that failure is virtually present in all expert performance and that learning and mastering know-hows is inevitably linked with mistakes and failures.

The second option involves, as we mentioned, rectifying CST so as to make room for occasional failure:

Counterfactual Success Thesis with occasional Failure [CSTF]:
x knows how to ϕ only if: if x tried to ϕ under normal conditions,
x would usually succeed in the relevant ϕ-ing[21].

We think that CSTFis a more accurate and defensible thesis: it is more accurate in the light of the last paragraph's remarks, as it leaves room for failure, and is more defensible for two reasons: first, it can be reinforced by well-established statistical methods, and second, as a consequence of using the statistical techniques, it leads to a couple of significant results with regard to knowledge-how and ability. Let us briefly mention two such results, one of which will be in our focus in due time. Firstly, the statistical construal of CSTF adequately addresses the problem of establishing thresholds[22] which certify the possession of knowledge how. Secondly, a certain desirable graduality in possessing an ability, and maybe know-how, emerges, but more on this later.

Returning to the first result, we note that equipping CSFT with more or less arbitrarily imposed thresholds that would establish the possession of knowhow or ability has at least two shortcomings that the statistical interpretation circumvents. The decision procedure of simply attributing ability for values exceeding the threshold lacks any other desirable quantitative qualifications, such as the probability of errors

---

[20] (Hawley, 2003, p. 24)

[21] (Hawley, 2003, p. 24)

[22] Hawley is definitely aware of the difficulty of using thresholds as means of determining the possession of knowledge how: 'Presumably, there is no exact threshold which qualifies the subject as knowing how to X under C. Again, the threshold may be set by reference to competent performers'(Hawley, 2003, p. 24).

(false positives and false negatives, or in statistical jargon, type I and type II errors); moreover, the attribution of ability or know how based on this procedure is insensitive to the number of trials, although it is intuitively obvious[23] that the number of trials should play a significant role.

In what follows, we are not going to fully articulate how statistical methods can be used to give credentials to CSTF; we will only sketch the idea behind how this can be achieved. And the rough idea is that we can use probability distributions and tests of statistical significance based on such distributions to specify how 'usually' in CSFT could be statistically construed:

Counterfactual Success Thesis Statistically construed [CSTS]:
x knows how to ϕ only if: if x tried to ϕ under normal conditions,
x would succeed in a statistically significant manner in the relevant ϕ-ing.

The statistical package to be used has one essential function: to give a mathematical meaning to the expression 'statistically significant' as it occurs in CSTS. We are going to cover how this is accomplished by means of a well-known example, but to give the headline, to establish whether the successes in the relevant ϕ-ing obtained by x in some trials are statistically significant means that they are (extremely) unlikely[24] to be caused by some stochastic process. With the discriminating tool of statistical testing deployed, we can infer, with a certain sub-unitary degree of confidence, of course, whether x's successes are likely due to some stochastic process or something else is likely at play, namely the expression of an ability. In this sense, we can say that results that fail to achieve statistical significance are more likely caused by randomness, or luck, or mere guesses, so they amount to lack of an appropriate ability, while those who achieve statistical significance are likely to be the expression of some sort of an ability.

Resuming, in order to establish whether x knows how to ϕ, according to CSTS, we need a mathematical framework that would enable us to infer whether the number of successes registered by x in successive trials of ϕ-ing qualifies x as succeeding in a statistically significant manner or not, that is, whether x has the corresponding relevant ability or not. As mentioned above, the inference has a statistical character, so, we need to qualify both the inference as well as the attribution of the ability/know-how based on statistical significance as having a (sub-unitary) degree of confidence.

[23] After all, there is a significant difference between the same percentage of successes out of, say, 10 trials or 1000 trials.

[24] How unlikely depends on a couple of factors which can be adjusted to various demands of precision.

In standard statistical tests of significance, this mathematical framework takes the form of a probability distribution of a variable[25] that quantitatively encodes the outcomes of a random process. Accordingly, an appropriate framework should provide a probabilistic description of the expected distribution of successes in all the accessible counterfactual situations, assuming the absence of the ability to ϕ – which, as we stated, corresponds to a random distribution of successes. Standard tests of statistical significance are built upon such an assumption, called the *null hypothesis*, hence the description of such tests as *null hypothesis significance testing* (NHST).In inferential statistics, where NHST are widely used, the standard procedure is to identify, assuming the null hypothesis, a mathematical model in the form of a probability distribution[26] (taking also into consideration the parameter and the characteristics of the population of interest, of course), against which the difference[27] between the observed and the expected value(s)is determined, together with the associated probability of observing such a difference. The probability distribution that follows from the null hypothesis and the relevant characteristics of the parameter & population of interest is described by the probability mass function (pmf)/probability density function (pdf)[28] or the cumulative distribution function (cdf). A decision procedure is, then, applied to the value of the test statistic, and, accordingly, the null hypothesis is rejected or not.

Next, we will show a demo of how the standard procedure in NHST can be successfully[29] applied to cases of know-how and corresponding relevant ability that form the subject matter of CSTS. To illustrate such an application, we are going to present a well-known experiment, *lady tasting tea*, first described and analyzed by Ronald A. Fisher[30]. A lady colleague of Fisher, Muriel Bristol, claimed to have the ability to distinguish whether the tea or the milk was added first to a cup. So, ϕ = distinguish whether the tea or the milk was poured first in a cup, and, in order to determine whether the lady would significantly succeed in the relevant ϕ-ing according to the NHST protocol, one needs to set up an adequate mathematical model. To this end, Fisher devised an experiment involving 8 cups of tea and milk, 4 of which were prepared by one method, the rest by the other. The lady would have to choose four cups prepared by the same method or recipe – tea first, milk

[25] Such variables are known as 'random variables'.

[26] The mathematical model can be a direct consequence of the null hypothesis, as will be the case with the example below, or a consequence of the sampling distribution of the parameter, thus, deriving from a more complex mathematical result, such as the Central Limit Theorem.

[27] The difference, measured in the standard deviations of the probability distribution, (and the associated probability – commonly known as p-value) represents the test statistic value(s).

[28] If the random variable is discrete, its distribution is described by pmf; pdf is used when the random variable is continuous.

[29] Pun intended.

[30] (Fisher, 1935)

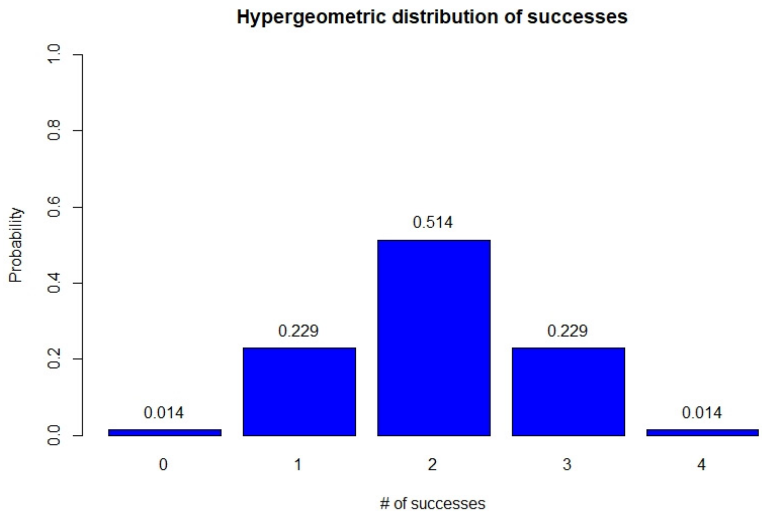
after, or milk first, tea after. Accordingly, the range of possible worlds determined by the design of the experiment is the sample space of all the possible outcomes, and the 'if x tried to ϕ under normal conditions, x would succeed in a statistically significant manner in the relevant ϕ-ing' part of CSTS is constituted by a statistically significant sequence of successes that resulted from the lady's choices. A simple combinatorial argument tells us that the number of different possible outcomes of choosing 4 cups out of a total of 8is: $\binom{8}{4}$ = 70.

In this setting, the null hypothesis, that the lady doesn't have the claimed ability – which amounts to considering that the number of successes is due to pure chance –, determines a well-known probability distribution of the number of successes: the hypergeometric distribution. Technically, the number of successes in this instance is a random variable $X \sim HG(n, K, N)$ whose pmf is:

$$P(X=x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{K}}, \text{ for } \max(0, n-N+K) \leq x \leq \min(n, K)$$

where n = # of draws, N = population volume, x = # of successes in the sequence of n draws, K = # of existent successes in N. The expected value of $X \sim HG(n, K, N)$ is $E(X)=\frac{nK}{N}$.

For the *lady tasting tea* case, $n = 4$, $N = 8$, $K = 4$; $x = \overline{0, 4}$,so E(X) = 2, and the probability distribution[31] of the number of successes is:



[31] We have plotted the probability distribution of the number of successes using 'R' software.

We can interpret[32] the probability histogram of the distribution of the number of successes in a modal setting, bearing in mind the connection between the vocabulary of modal logic and that of statistics: the range of possible worlds is the sample space determined by the design of the experiment, the (hypergeometric) distribution of successes determines a partition on the sample space (so on the range of possible worlds), and the probability associated with each possible number of successes is the weight of the corresponding equivalence class of possible worlds relative to the range of all such possible worlds. In this setting, we can interpret the probability histogram as a description of the accessibility relation equipped with a probability or weight that follows from centering on a world that is characterized by the null hypothesis.

Now, as can be seen from the probability distribution histogram, choosing all 4 cups prepared by the same method is so far form the expected value E(X) of choosing only 2 such cups, that the probability of observing such a result assuming – and we cannot underemphasize the importance of this assumption – the null hypothesis, that the lady doesn't have the corresponding ability, is approximately 1.4%. It should be obvious that the all-draws-successful probability is1/70, that is, the number of favorable cases corresponding to all-draws-successful divided by the total number of possible cases, i.e. the cardinality of the sample space.

Fisher argued, based on a statistical norm active among his peers, that probabilities of observing the data under 5%, assuming the null hypothesis, should be interpreted as a good indication that the null hypothesis should be abandoned. To this day, this decision rule is active in interpreting statistical results in social sciences[33]. In our case, the only acceptable case for rejecting the null hypothesis, according to this decision rule, would be for the lady to correctly identify all the 4 cups prepared by the same method. Anecdotal evidence[34] suggests that the lady managed to have done that, so the conclusion of Fisher's experiment was that the lady, most likely, possessed such a know-how or ability. In NHST, the decision rule is

[32] The interpretation is not original, it just builds upon an analogy that Kripke draws between his conception of modal logic and the way probabilities are conceived in the mathematical framework of probability theory(Kripke, 1980, pp. 16-18).

[33] Of course, there are other, more severe, thresholds in social sciences; in physics, for example, as it employs statistical methods, the standard for rejecting the null is extremely higher than in social sciences, but that is partly due to the difference in variability of the phenomena studied by the two fields.

[34] David Salzburg, in his 2001 book, *The Lady Tasting Tea. How Statistics Revolutionized Science in theTwentieth Century* writes that 'Fisher does not describe the outcome of the experiment that sunny summer afternoon in Cambridge. But Professor Smith (H. Fairfield Smith, a colleague of Fisher,– my note) told me that the lady identified every single one of the cups correctly'(Salzburg, 2001, p. 8).

usually more complex, requiring a precise specification of the critical values for the rejection area of the probability distribution in question, according to the type of test to be performed (one-tail, two-tails), but the underlying idea is that the critical values on the probability distribution should correspond to a p-value of at most 5%.

Now, what we have just described is an example of a real-life application of NHST protocols to determine whether a certain individual has a certain ability; obviously, the statistical inference that the lady, most likely, has the claimed ability is probabilistic (as we mentioned several times above); we didn't establish with certainty that the lady possessed such an ability, only that the lack of such an ability on her side is improbable.

Obviously, Fisher's exact test, as it is known, is not a singleton on the statistical market of tests, and the adequate way of testing statistical significance depends on the ϕ in question, on the characteristics of the population, etc., more generally, on the particular situation that is analyzed. Of course, there certainly are cases that elude such a straightforward modelling, for example, because there isn't a clear-cut way to define success. Various solutions can be found and implemented in such cases, all heavily dependent on the particularities of the situation under analysis, but a proper analysis of such solutions is too lengthy to be addressed in this paper. Here, we limit ourselves to enumerate a couple of easily accessible such solutions: maybe a threshold to define success could be consensually agreed upon, maybe other statistical methods could be used to transform performances measured on a non-dichotomic scale into a dichotomous one without losing any relevant information, maybe new tests can be developed in order to answer such demands, we certainly don't intend to assert that such solutions are always available, so that, in principle, could exhaust all the cases that constitute the subject matter of CSTF.

The statistical tests used to discern the likelihood of the possession of an ability ϕ, can be, and, in statistical practice, are doubled by considerations regarding the effect size. For our purposes, it suffices to say that the effect size measures[35] the magnitude of the difference between the observed and the expected values, so, it is a reliable evaluation of the strength or the intensity of the results usually obtained in a previous step of statistical significance testing. In a sense, we can say that the effect size informs us about the importance or impact of the observed difference. The distinction between tests of statistical significance and measures of effect size can be sharpened by looking at the interpretations attached to these procedures and their corresponding values: a statistical significance test tells us whether *there is* an effect, under the form of a statistically significant difference, whereas the value of the effect

[35] The measurement is usually in appropriate (i.e. contextually-determined) standard deviations, but see the subsequent discussion.

size *measures the magnitude of this effect*, whether it is small, medium, strong, for example. So, it should come as no surprise that one can encounter results that are statistically significant, yet their effect size is small, so, although there is a significant deviation from the expected value (under the null hypothesis), the deviation in question is small, and, often, unimportant in practice. To illustrate further, suppose that under the null hypothesis we expect a mean value of 15, we observe a mean value of 17, that differs statistically significant from the expected, but the effect size, measured by Cohen's d[36], say, is 0.1. Then, under the widely used guide of interpreting indicators of effect size proposed by Cohen[37], the above-mentioned effect size is considered to be small[38], so, the difference, we can safely conclude, is not important. As the illustration suggests, the effect size is not measured by a one-size-fits-all measure, there are numerous custom-made indicators to adequately account for the different parameters and populations of interest; a discussion about the complexities involved in the design and use of adequate measures of effect sizes is a vast subject, way beyond the scope of this paper. Before turning to some philosophically relevant consequences of using the statistical package of NHST & effect size, it is worth mentioning that because CSTS is expressed in terms of success and failure, the distributions involved in the statistical procedures of testing the possession of an ability are, mathematically speaking, simple, mostly binomial and hypergeometric, but that doesn't mean the there is an accompanying mathematically simple and easy to understand measure of the effect size. In fact, several indicators for measuring effect size have been proposed for the distributions relevant to CSTS, ranging from relative risk (RR), odds ratio (OR), risk difference (RD), to the binomial effect size display (BESD), number needed to treat (NNT)or Cohen's h[39]. In the *lady testing tea* case, several reliable measures not included in the list above could be used to determine the effect size, by displaying the data (and consequently construe Fisher's exact test) in a 2x2 contingency table and use statistical indices specific for such tables for measuring the strength of an association, such as Pearson's ϕ, Cramer's V, the contingency coefficient, etc.

The variability of the size effect implies that the ability tested according to CSTS varies also, so, in a sense, a consequence of interpreting CST statistically is that ability should be placed on a scale, not considered a have-it-or-not attribute.

---

[36] Cohen's d is a popular measure of the effect size, for its formula and use see (Cohen, 2008).

[37] (Cohen, 2008)

[38] Because Cohen's d is measured on the standard deviation metric of an approximately normal curve, the value 0.1 could be interpreted as indicating that the percentage of values that are below the observed mean of 17 is 53.98%, under the assumption of the null hypothesis.

[39] See *https://www.psychometrica.de/effect_size.html* for more information, examples and online calculators of these indices.

This graduality of ability is certainly something desirable. It accounts, first, for the difference we see in the possession of the same ability: we often describe someone as being beginner or intermediate or advanced with regard to some ability. Secondly, it accounts for the evolution of someone who is in the process of mastering a certain ability. The fact that a certain performance significantly deviates from what we expect under the assumption of no ability, but the effect size is small, can be interpreted as a description of someone who is in the process of mastering the ability in question. In the same vein, we can view and measure the difference between the performance of an expert and that of a beginner, by different values of the same measure of the effect size. This graduality of ability leaves us with an interesting ‘contraposition’ of the anti-intellectualist thesis that knowledge how implies ability or counterfactual success: that the graduality of ability implies the graduality of knowledge how. We are not stating here that anti-intellectualists have to embrace the graduality of knowledge-how as a consequence of our statistical interpretation, just that, under suitable conditions, an anti-intellectualist could advance the thesis that know-how, at least in some instances, is not a ready-made-clear-cut concept whose possession one either has it or doesn’t have, but something more fluid and with a fleeing texture, that individuals, through practice, gradually acquire. In a sense, we are saying that through practice, a tennis player gradually develops not only their skill and ability, but also gradually acquires know-how; after all, there isn’t a definitive moment at which we can categorically affirm the player is now in full possession of the appropriate knowledge how. Of course, this is a controversial claim, that needs further theoretical analysis and experimental testing.

REFERENCES

Bengson, J., & Moffett, M. A. (2012). Nonpropositional Intellectualism. In J. Bengson, & M. A. Moffett (Eds.), *Knowing How. Essays on Knowledge, Mind, and Action* (pp. 161-196). New York: Oxford University Press.

Cath, Y. (2015). Revisionary intellectualism and Gettier. *Philos Stud, 172*, 7-27. doi:https://doi.org/10.1007/s11098-013-0263-y

Cohen, J. (2008). *Statistical power analysis for the behavioral sciences* (Second ed.). Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc.

Fantl, J. (2017). Knowledge How. *The Stanford Encyclopedia of Philosophy*. Retrieved from https://plato.stanford.edu/archives/fall2017/entries/knowledge-how/

Fisher, R. A. (1935). *The Design of Experiments.* Edinburgh: Oliver and Boyd.
Hawley, K. (2003, January). Success and Knowledge-How. *American Philosophical Quarterly, 40*(1), 19-31. Retrieved from http://www.jstor.org/stable/20010094
Kripke, S. (1980). *Naming and necessity.* Cambridge, Massachusetts: Harvard University Press.
Ryle, G. (1949). *The concept of mind.* London: Hutchinson.
Salzburg, D. (2001). *The Lady Tasting Tea. How Statistics Revolutionized Science in the Twentieth Century.* New York: W.H. Freeman.