A Fuzzy Take on the Logical Issues of Statistical Hypothesis Testing

Matthew Booth and Fabien Paillusson *

School of Mathematics and Physics, University of Lincoln, Lincoln LN6 7TS, UK; mbooth@lincoln.ac.uk
* Correspondence: fpaillusson@lincoln.ac.uk

Abstract: Statistical Hypothesis Testing (SHT) is a class of inference methods whereby one makes use of empirical data to test a hypothesis and often emit a judgment about whether to reject it or not. In this paper, we focus on the logical aspect of this strategy, which is largely independent of the adopted school of thought, at least within the various frequentist approaches. We identify SHT as taking the form of an unsound argument from Modus Tollens in classical logic, and, in order to rescue SHT from this difficulty, we propose that it can instead be grounded in t-norm based fuzzy logics. We reformulate the frequentists' SHT logic by making use of a fuzzy extension of Modus Tollens to develop a model of truth valuation for its premises. Importantly, we show that it is possible to preserve the soundness of Modus Tollens by exploring the various conventions involved with constructing fuzzy negations and fuzzy implications (namely, the S and R conventions). We find that under the S convention, it is possible to conduct the Modus Tollens inference argument using Zadeh's compositional extension and any possible t-norm. Under the R convention we find that this is not necessarily the case, but that by mixing R-implication with S-negation we can salvage the product t-norm, for example. In conclusion, we have shown that fuzzy logic is a legitimate framework to discuss and address the difficulties plaguing frequentist interpretations of SHT.

Keywords: logic; fuzzy logic; probability; decision making; philosophy of science

1. Introduction

According to Popper [1], a theory can be deemed scientific if it implies the *impossibility* of certain events that are otherwise perfectly conceivable (or even expected). A test of a scientific theory should thereby be an attempt at falsification by means of observing such impossible events. To many observers this refutability strategy was an important step towards solving the induction problem that plagues the philosophy of science and, at the same, undermines the legitimacy of scientific claims. While the Popperian view on scientific practice is far from being the final say on what constitutes a scientific theory [2], it is nevertheless the most popular and invoked philosophy of science in scientific circles. From a logical standpoint the rejection argument for a hypothesis H given evidence E operates via the *Modus Tollens* (MT) inference rule:

$$H \implies \neg E \tag{1}$$

$$E \tag{2}$$

$$\therefore \quad \neg H \tag{3}$$

where $\neg$ represents the negation symbol and $\implies$ an entailment. According to this logic, a solitary observation of a falsifying event E necessarily undermines the hypothesis H. In practice, however, since there is no such thing as a 'pure observation', a single counter-instance does not falsify the theory. There are at least two reasons why this is the case. First, the research hypothesis H being tested is necessarily supplemented by a set of

auxilliary hypotheses H' (i.e., underlying 'hidden' assumptions) that connect it to the real world. What is actually being tested, therefore, is the conjunction $H \wedge H'$. This is an example of holistic underdetermination; one can falsify this conjunction but not the research hypothesis alone [3]. Second, experimental data is always susceptible to statistical uncertainties and it is unclear in what sense MT can still hold in such cases. This question will be the central theme of the present paper. That evidence is subject to statistical effects was already known and appreciated in the last decades of the 19th century as an integral part of scientific practice. However, this was not addressed quantitatively until the beginning of the 20th century with Student's exact derivation of the *t*-distribution in probability theory, and was turned into a more general framework by Fisher, Neynman and Pearson in the 1920s [4]. In what follows we shall refer to this alleged general framework as *Statistical Hypothesis Testing* (SHT). To prevent any misunderstanding, we wish to clarify that we use SHT here as an umbrella term for the various classical, frequentist methodologies and philosophies pertaining to hypothesis testing under statistical considerations. As a result, the present study does not address the developments in hypothesis testing that rely on Bayesian inference. The interested reader may explore some of the tools, challenges and promising avenues of Bayesian hypothesis testing in Refs. [5–8]. The main frequentist schools we shall consider under the SHT term are Fisher's [9], Neyman and Pearson's [10] and the *Null Hypothesis Significance Testing* (NHST) school; the latter often being considered as the illegitimate child of the two former [11]. Despite their very strong philosophical differences, it is our appreciation that these three schools of thought can be summarised by a reasoning of the form

$$\text{P1}: \quad H \implies (P(E|H) \geq \alpha) \tag{4}$$

$$\text{P2}: \quad \neg(P(E|H) \geq \alpha) \tag{5}$$

$$\therefore \qquad \overline{\qquad \neg H \qquad} \tag{6}$$

In the above argument, $P(E|H)$ is the a priori knowable conditional probability of some observed evidence E given an assumed true hypothesis H. P1 should therefore be interpreted as a proposition about E: P1 $\equiv$ "If H then E is such that $P(E|H) > \alpha$". Likewise P2 is also about the observed evidence E: P2 $\equiv$ "E is such that $P(E|H) < \alpha$". The quantity $P(E|H)$ is the so-called P value of the observed evidence and is a commonly used frequentist measure in medical studies and life sciences. Different schools attribute different meaning and value to the reasoning in Equations (4)–(6). For example, 'Fisherians' would hold that a hypothesis can eventually be intellectually rejected under certain circumstances via an inductive reasoning while 'Neymanians' would contend that the hypothesis itself can usually not be rejected but a research worker may behave *as if* it was rejected [12]. Of the three schools we are discussing here, NHST is by far the most predominant methodology being used nowadays and as such has received many critics on two principal accounts: (a) its methodological framework comes as an incomplete and simplified mix of Fisher's and Neyman and Pearson's philosophies which does not take into account the more subtle but necessary attributes of either of these respective schools [11]. (b) As a result, NHST practitioners often misunderstand and mischaracterise the meaning of the numbers they obtain through the application of NHST's various statistical procedures [13,14]. A commonly held false claim from a substantial portion of NHST practitioners is that their methods enable them to know the probability for the chosen hypothesis to be false [14], i.e., from the reasoning in Equations (4)–(6) some NHST practitioners would claim that they know $P(H|E)$. As pointed out by Goodman [15], this is of course fallacious. From Bayes' rule we have the following relationship between $P(H|E)$ and $P(E|H)$:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|\neg H)(1 - P(H))} \tag{7}$$

where both the prior $P(H)$ and the conditional probability $P(E|\neg H)$ for E to be observed given H being false have unknown values. Goodman contends then that:

1. Evaluating how likely is H given the observed evidence E requires knowing $P(H|E)$,
2. Evaluating $P(H|E)$ requires the knowledge of two additional variables $P(H)$ and $P(E|\neg H)$ which are independent from $P(E|H)$,
3. It is therefore not possible to reject H on the sole basis of observing E and knowing a priori $P(E|H)$.

This conclusion can be illustrated by plotting the probability density map of $P(H|E)$ for $P(E|H) = 0.04$ (i.e., usually considered as statistically significant in medical statistics) with $P(H)$ and $P(E|\neg H)$ taking all possible values between 0 and 1.

In Figure 1, we see that despite the fact that $P(E|H) = 0.04$, about half of the parameter space (most of the bottom right corner) would give more than 20% chance H to be true given E. In this light, it becomes inevitable to see hypothesis rejection on the sole basis of P value as an invalid, fallacious argument. This problem—that what could be inferred from a reasoning of the type (4)–(6) was not quantified by an actual probability—was already known and pondered by pioneers of SHT. Fisher [4] would admit that:

> *...more generally, however, a mathematical quantity of a different kind, which I have termed mathematical likelihood, appears to take its place [i.e., the place of probability] as a measure of rational belief when we are reasoning from the sample to the population.*

When the hypothesis H can be specified as one or more parameter values in a model then the likelihood $\mathcal{L}(H|E) \equiv P(E|H)$ is the quantity that seems to be recommended for rational reasoning here. Neyman and Pearson [10] shared a similar opinion

> *Here and later the term "probability" used in connection with Hypothesis A [i.e., that a sample data is a finite realisation of a theoretical population] must be taken in a very wide sense. It cannot necessarily be described by a single numerical measure of inverse probability; as the hypothesis becomes "less probable" our confidence in it decreases, and the reason for it lies in the meaning of the particular contour system that has been chosen [i.e., critical regions].*
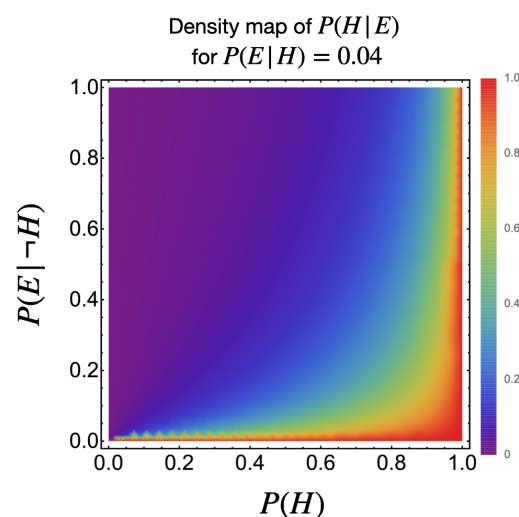

Figure 1. Density map plot of $P(H|E)$ for $P(E|H) = 0.04$ for $P(H)$ and $P(E|\neg H)$ taking independently all possible values between 0 and 1. Colours leaning towards purple refer to low probability values while redder colours correspond to probabilities close to 1.

In the above citation, texts in square brackets were added to provide missing context. This weakened version of probability to make a decision about H is contrary to Bayesian inference whereby the only thing newly obtained data can permit is to update a pre-existing belief about H quantified by $P(H)$. Furthermore, this prior is not enough for the update to

occur as indicated in Figure 1. This constitutes a traditional critique of Bayesian approaches in that they are claimed to rely on the subjectivity of the practitioner. Such a blanket critique of Bayesian strategies on the sole basis that prior distributions are needed is, however, unwarranted on two grounds. First, it is illusory to think that non-Bayesian analyses escape subjectivity since they always require individual expert knowledge to set the hypotheses, alternative hypotheses and α-levels of significance. Second, it dismisses a fruitful research programme in Bayesian analysis that seeks to develop objective priors by generalising and formalising notions such as the principle of indifference and the principle of insufficient reason [16]. How these could be used in the case of Bayesian hypothesis testing is discussed for example in [7]. In principle, the statistician is thus faced with a conundrum. They can either use a frequentist approach and be criticised on the grounds that their reasoning can never give the—assumed necessary—probability for the hypothesis to be false given the evidence, or they may use a Bayesian approach and, other issues aside, be asked to justify its objectivity. The present paper is concerned only with the former case, since this is the more heavily used practice in contemporary research. Thus, even if one assumes that the likelihood, or some wider notion of probability, may be key to expressing a valid pseudo-inductive judgment, the question we wish to address in the present study remains: does the reasoning evoked in Equations (4)–(6) hold-up to that promise? If one works with a logic where MT holds, then this argument is valid by construction. Whether the argument is sound or not, i.e., whether both P1 and P2 are true or not, is more questionable. If we are forced to classify P1 as being either true or false then it is must necessarily be false since, under *H* being true, any *E* with non-zero probability can in principle be observed, not only those with a probability higher than a set threshold. Thus, from a classical logic point of view, basing the rejection of a hypothesis solely on the *P* value is also fallacious. That there cannot exist a Probabilistic Modus Tollens was also put forward by Sober [17] and Greco [12] whom essentially echo a sentiment by Neyman and Pearson [10]

> *It is indeed obvious, upon a little consideration, that the mere fact that a particular sample may be expected to occur very rarely in sampling from* Π *[i.e., the hypothetical population] would not in itself justify the rejection of the hypothesis that it had be so drawn, if there were no more probable hypotheses conceivable.*

This line of argument in the philosophy of SHT [10,12,17] considers that the MT reasoning developed in Equations (4)–(6) is basically fallacious but can be rectified by considering *H* as being rejected in favour of an alternative, more likely hypothesis, rather than being rejected as a standalone hypothesis. While this strategy is clearly trying to infer the most out of the available data [18], we believe that it does not permit SHT as a reasoning strategy to be grounded in classical logic. More specifically, one cannot claim that the reasoning in Equations (4)–(6) is fallacious and state at the same time that a judgement based on likelihood ratios is valid without contradicting themselves. The reason for this is that the likelihood ratio is still subject to statistical effects and one still needs to evaluate the corresponding probability for this ratio to lie in a certain critical region; which is equivalent to the reasoning in Equations (4)–(6). As Neyman and Pearson put it [10]

> *It would be possible to use the [likelihood] ratio* λ *as a criterion but this without the knowledge of* P_λ *does not enable us to estimate the extent of the form (1) error.*

The form (1) error mentioned being the frequency rate with which one may wrongly reject the tested hypothesis upon repeating many times the same experiment. ideally, one would want this form (1) error to be as small as possible (usually set by the value α in Equation (4)) under other specific constraints set by the experimenter (e.g., minimizing other forms of judgement error).

Based on the above discussion we agree with Sober [17] that "there is no probabilistic analog of modus tollens" but we have also shown that a) this fundamental state of affairs could not be salvaged by simply comparing hypotheses and b) the logical problem has been shown to occur when propositions need to be exactly true or false. We cannot deduce

from this restricted analysis that there is no logic within which SHT can make sense. Indeed, in the present work, we explore the possibility of grounding SHT—as described in Equations (4)–(6)—in fuzzy logic rather than in conventional classical logic. Section 2 is dedicated to present and motivate many-valued logics fuzzy logics. Section 3 discusses how certain formulations of fuzzy logic might enable SHT to be formulated as a valid and sound argument. Finally, we evaluate more closely which of the various conventions surrounding fuzzy negation and fuzzy implication provide a viable logic with which to ground frequentist SHT and discuss where do our findings leave us with regards to scientific reasoning in Section 4.

2. Classical Logic, Multivalued Logic and Fuzzy Logic

Classical logic or propositional logic is a formal language which comprises: a set of atomic propositions $D = (a, b, c, ...)$; a set of connectives $C = (\neg, \vee, \wedge, \Longrightarrow)$, where $\neg$ stands for *negation*, $\vee$ for *disjunction*, $\wedge$ for *conjunction* and $\Longrightarrow$ for *implication*, that bind multiple atomic propositions into sentences; and a valuation function $v : C \mapsto \{0, 1\}$, which assigns a truth value of either 1 or 0 to the connectives depending on the truth value of the propositions they bind (cf. Table 1). As we have seen in Section 1, it is this last feature that creates a problem for SHT; the MT inference rule in classical logic yields perfectly true and exact conclusions only to the extent that the premises themselves are perfectly true and exact. Yet, as we have discussed, the premises involved in the P value reasoning from Equations (4)–(6) are necessarily false. For example, if $\alpha = 0.05$ in premise P1 and one interprets E as $S > s$ (where S is some aggregate random variable and s its realisation), then one can conceive of realisations s (i.e., models) where H is true but $P(S > s|H) < 0.05$, thereby automatically making P1 false.

Table 1. Truth table of classical logic.

a	*b*	$\neg a$	$a \wedge b$	$a \vee b$	$a \Longrightarrow b$
0	0	1	0	0	1
0	1	1	0	1	1
1	0	0	0	1	0
1	1	0	1	1	1

What if, however, one departs from the idea that propositions must necessarily be either true or false by adopting shades of truth? Allowing such a middle ground between true and false to exist might help model the fact that one may credit a given proposition as being, for example, "more true than false" (albeit not strictly true). The desire to apply logic to situations where propositions are perceived as partially true (or indeed where some things are "truer" than other things [19]) has motivated the development of many-valued logics (MVLs)—logics which accommodate truth values that are intermediate between true and false. The concept of MVL is far from new and can be traced at least as far back as the 4-valued logics used in Buddhist and Greek philosophies under the names of *Catuskoti* and *Tetralemma*, respectively, [20]. Of course, adopting a number-based representation of truth values enables a greater diversity of MVLs. The principle of bivalence that characterises classical logic is rejected within an MVL, and the valuation function v may thus assign a truth value that is between 0 and 1. The truth value of a sentence is still contingent upon the truth values of the propositions that the connectives bind, but the degree of truth associated with the atomic propositions themselves are no longer two-valued.

Infinitely many-valued logics (IMVLs) are a subclass of MVLs in which the valuation function maps to a continuous variable on the interval $[0, 1]$, such that there are infinitely many truth values that a proposition could be assigned. The most widely used IMVLs rely on so-called t-norms $\top(x, y)$ between two numbers $x, y \in [0, 1]$. A t-norm is a binary operator which has the following properties:

1. $\top(x, y) = \top(y, x)$;
2. $\top(x, y) \leq \top(w, z)$ if $x \leq w$ and $y \leq z$;
3. $\top(x, \top(y, z)) = \top(\top(x, y), z)$; and
4. $\top(x, 1) = x$.

In Table 2, the third column represents the conjunction operator whose valuation is identified to a t-norm in all t-norm based IMVLs. Columns 4-6 show how the conditional, negation and disjunction connectives are defined under the S-implications convention: a strong negation is defined as $\nu(\neg_S) \equiv 1 - \nu(a))$, the corresponding S-implication is defined from $a \implies_S b \equiv (\neg_S a \vee_S b)$, and the S-disjunction is defined via the t-conorm $\bot_S(x, y) \equiv \top(1 - x, 1 - y)$. Columns 7–9 introduce the negation, disjunction and conditional connectives under the *residuum* or R-conventions: the R-implication is defined as the residuum of the IMVL adopted t-norm via $\nu(a \implies_R b) \equiv \max\{z | \top(z, \nu(a)) \leq \nu(b)\}$ and the R-negation is defined from $\neg_R a \equiv a \implies_R 0$, where the zero-valuation on the right-hand side of the implication refers to a proposition false in all models. Finally, the R-disjunction is defined with respect to the corresponding t-conorm $\nu(a \vee_R b) \equiv \bot_R(x, y) \equiv \top(\nu(\neg_R a), \nu(\neg_R b))$. Note that R-implications and R-disjunctions may be used with negations other than the R-negation [21].

There is a certain degree of freedom in how one goes about choosing a given t-norm. Rows 3–5 of Table 2 give the explicit expressions for all the connectives for three common t-norms: Godel, product and Lukasiewicz.

Interestingly, granting a truth value in the interval $[0, 1]$ to logical propositions has led to further developments in set theory, pioneered by Zadeh [22]. The idea consists first in adopting a representation of traditional sets as follows. Consider a subset A of a larger measurable set Ω. The set A can be characterised by its characteristic function χ_A acting on Ω such that for any $x \in \Omega$, $\chi_A(x) = 1$ if $x \in A$ and $\chi_A(x) = 0$ otherwise. The set A can be represented as a realisation $\chi_A \in \{0, 1\}^\Omega$ which assigns a value 0 or 1 to all elements Ω. Zadeh then introduced the notion of *fuzzy sets* for which a given element in Ω can only partially belong to the set A with a membership function μ_A returning a value between 0 and 1. In analogy to the previously introduced notation, fuzzy sets are then representable as a realisation $\mu_A \in [0, 1]^\Omega$. On this view, traditional sets are then called *crisp sets* since the membership function can only return either 0 or 1. Typical operations between fuzzy sets such as complement, intersection, union and containment are then defined from their IMVL counterparts, i.e., t-norms, t-conorms and negation. In this way, fuzzy set theory enables vague linguistic variables such as short, tall, large and small, to be represented as fuzzy sets, so that inferences may be made about the relations between them. In its narrowest definition [23,24], a *fuzzy logic* is a logic that applies to fuzzy statements expressed as fuzzy sets, such that fuzzy inferences can be made. As such, fuzzy logic can be construed as an extension of IMVL with some additional structures that enable the handling of fuzzy notions. In what follows we will use *t-norm based fuzzy logics*, which are grounded in t-norm based IMVL and supplemented by some fuzzy rules. In Section 3, we shall rely on a specific fuzzy inference rule while in Section 4 we shall briefly discuss the advantage that fuzzy variables may bring to the problem we are attempting to frame.

Table 2. Truth tables of t-norm-based IMVLs for the three most common t-norms.

$a \longrightarrow b$		$a \wedge b$	$a \vee_S b$	$\neg_S a$	$a \Longrightarrow_S b$	$a \vee_R b$	$\neg_R a$	$a \Longrightarrow_R b$
x	y	$\top(x,y)$	$\bot_S(x,y)$	$1-x$	$\bot(1-x,y)$	$\bot_R(x,y)$	$\max\{z\mid \top(z,x) \leq 0\}$	$\max\{z\mid \top(z,x) \leq y\}$
Godel		$\min(x,y)$	$\max(x,y)$	$1-x$	$\max(1-x,y)$	1 if $xy>0$, 0 if $x=y=0$	0 if $x>0$, 1 if $x=0$	$\min(1,y)$
Product		xy	$x+y-xy$	$1-x$	$1-x+xy$	1 if $xy>0$, 0 if x=y=0	0 if $x>0$, 1 if $x=0$	$\min(1,y/x)$
Lukasiewicz		$\max(x+y-1,0)$	$\min(x+y,1)$	$1-x$	$\min(1-x+y,1)$	$\min(x+y,1)$	$1-x$	$\min(1-x+y,1)$

3. Application of Fuzzy Logics to Hypothesis Rejection

In the previous section we have introduced tools with which it is possible to reason about partial truths. Within the generic MT argument for hypothesis rejection presented in Equations (4)–(6), we can then try to assign a truth value to the premises P1 and P2. Before doing so however, it is worth illustrating how probabilities and truth values differ from each other as a matter of principle. Probabilities, in their common acception, refer to the chances that propositions are exactly true. If we say that "the probability for Tom to be tall is 0.3", then it usually means that we do not know Tom's height but that he belongs to an equivalence class of people among which 30% are tall. If, on the other hand, one states that "the truth value for Tom to be tall is 0.3" then it means that Tom, as an individual, is only marginally tall. Notwithstanding this difference between truth valuations and probabilities, it is still the case that the former can be informed by the latter depending on the situation. In fact, if we interpret a truth value as a degree of confidence in a given proposition, then it can surely be influenced by probabilistic considerations if the said proposition has a statistical connotation. It is our understanding that P1 carries a statistical meaning by asserting that, given H, the probability with which evidence E is observed is larger than some value α. This is because, to some extent, P1 can be interpreted as a proposition about the future and, following Carl Friedrich von Weizsacker, the future is deeply connected to probabilities [25]. Of course, if $\alpha = 0$ then P1 is a tautology because the consequent $P(E|H) \geq 0$ is always true. On that basis, it is tempting to suggest that the truth value of P1 is bound from below by the probability of the consequent to occur. Assuming the observed evidence is sampled from the conditional probability measure, this probability is $1 - \alpha$ (this can be seen in the case of observed evidence taking the form $E \equiv S \geq s$. In that case $P(S \geq s|H)$ is connected to a cumulative distribution and $P(S \geq s|H) \geq \alpha$ is a condition on it. Given that a cumulative distribution is monotonous, it can be translated onto a condition for s as $s \geq P^{-1}(\alpha|H)$. Assuming a probability measure $\mu_H(s)$ over s we get that the measure of the set of points s for which the condition is satisfied is $\int_{P^{-1}(\alpha|H)}^{+\infty} d\mu_H(s) = \int_0^{+\infty} d\mu_H(s) - \int_0^{P^{-1}(\alpha|H)} d\mu_H(s) = 1 - P(P^{-1}(\alpha|H)|H) = 1 - \alpha)$. Therefore, we may posit that the truth value for P1 satisfies $\nu(\text{P1}) \geq 1 - \alpha$. Note that upon conservatively choosing $\nu(\text{P1}) = 1 - \alpha$ for the valuation of P1, we do obtain truth values different from unity whenever $\alpha > 0$: the difference is now that there is no direct collapse into complete falsity. This intuition can be made more formal when trying to assign a probability to a conditional statement. Nguyen et al. [26] have proposed an interpolating expression between logical truth valuation function (which they call logical probability) and statistical probability for which the minimum valuation for a conditional $\nu(A \Longleftarrow B) = P(A|B)$, which is equal to $P(E > P^{-1}(\alpha|H)|H) = 1 - \alpha$ in our case (which agrees with our previous reasoning).

Contrary to P1 the valuation for P2 is fairly straightforward since it concerns actually observed evidence. So, given an a priori probability $P(E|H)$ and an observed evidence E (which we consider infinitely accurate for now) such that $P(E|H) < \alpha$, there is no particular doubt about the truthiness of the premise P2. As a result we assign a truth value $\nu(\text{P2}) = 1$ to P2. Note again the striking difference with the statistical interpretation: if P2 were to be interpreted statistically, there would usually be some very low probability to observe some evidence with conditional probability lower than α. In fuzzy logic, however, we instead assign a truth value to an actual observation E and a mathematical outcome $P(E|H) < \alpha$ following this observation; hence the truth value of 1. Finally, it is important to stress that $1 - \alpha$ has been assigned to P1 and *not* to the consequent $P(E|H) \geq \alpha$. Indeed, for the overall argument to be consistent, the truth value of $P(E|H) \geq \alpha$ has to be connected to our choice that $\nu(\text{P2}) = 1$.

To finish discussing how MT can be used within fuzzy logic, we need to appreciate that because the premises can in principle have any value between 0 and 1, so does the conclusion. However, since MT is a rule and not a connective, we cannot assign a value to this conclusion from Table 2 alone. Instead, we need to posit an extension of MT which preserves some of the desirable properties of such an inference rule [27]. In what follows

we will use Zadeh's *compositional extension* which, for MT, prescribes that the truth value of the conclusion $\neg H$ is

$$\nu(\neg H) \equiv \top(\nu(\neg[P(E|H) \geq \alpha] \implies \neg H), \nu(P(E|H) < \alpha)). \tag{8}$$

To evaluate Equation (8) we need to distinguish two possible cases: whether one is dealing with S-Implications or R-Implications. The reason lies in the fact that the first argument of the t-norm in Equation (8) is the valuation of the sentence $\neg[P(E|H) \geq \alpha] \implies \neg H$ which is the contrapositive of P1. Depending on the kind of implication being used, extracting $\nu(\neg[P(E|H) \geq \alpha] \implies \neg H)$ from $\nu(\text{P1})$ is not trivial.

3.1. Modus Tollens Under S-Implications in Fuzzy Logic

Under an S-Implication with a provided (canonical) negation $\nu(\neg_S a) \equiv 1 - \nu(a)$ we get that for any proposition a and b we have:

$$
\begin{aligned}
\nu(\neg_S b \implies_S \neg_S a) &\equiv \bot(\nu(\neg_S \neg_S b), \nu(\neg_S a)) = \bot(\nu(b), \nu(\neg_S a)) && (9)\\
&= \bot(\nu(b), 1 - \nu(a)) = \bot(\nu(b), 1) - \bot(\nu(b), \nu(a)) = \bot(1, \nu(b)) - \bot(\nu(a), \nu(b)) && (10)\\
&= \bot(\nu(\neg_S a), \nu(b)) \equiv \nu(a \implies_S b) && (11)
\end{aligned}
$$

where we used the fact that $\neg_S \neg_S a \equiv a$ in Equation (9) and both the distributivity and permutation symmetry of the t-conorm were used in Equation (10).

Equation (11) is called *contrapositive symmetry* and shows that it holds for any S-Implication. From this contrapositive symmetry we get that Equation (8) reads

$$\nu(\neg_S H) = \top(\nu(\text{P1}), \nu(\text{P2})) = \top(1 - \alpha, 1) = 1 - \alpha \tag{12}$$

for all t-norms. For the last equality of Equation (12) we used axiom 4 of t-norms which identifies 1 as the neutral element.

At this stage, it is worth checking whether the MT inference derived above is internally consistent, *i.e.* whether or not the conclusion contradicts some of the premises. One can see quite quickly that it is possible to assign a non-zero truth value to an S-Implication even if the consequent is exactly false (which is the case here because $\nu(\text{P2}) = 1$). In fact for the three models of t-norms we have chosen (i.e., Godel, product and Lukasiewicz) the outcome of the S-Implication in P1 for a false consequent is $1 - \nu(H)$. Since we have assigned $\nu(\text{P1}) = 1 - \alpha$, it follows that $\nu(H) = \alpha$. Under the negation rule we have used we get $\nu(\neg_S H) = 1 - \nu(H) = 1 - \alpha$; which is consistent with the final result of Equation (12).

3.2. Modus Tollens Under R-Implications and R-Negations in Fuzzy Logic

For R-Implications we are going to see that the consistency of the proposed argument depends on the chosen t-norm.

- Lukasiewicz t-norm: as can be seen from Table 2, the R-negation for this t-norm is the canonical negation $\nu(\neg_R a) = 1 - \nu(a)$. Applying this to the truth value of the contrapositive to an implication of the form $a \implies_R b$ we get

$$
\begin{aligned}
\nu(\neg_R b \implies_R \neg_R a) &\equiv \min(1, 1 - \nu(\neg_R b) + \nu(\neg_R a)) && (13)\\
&= \min(1, 1 - (1 - \nu(b)) + (1 - \nu(a))) = \min(1, 1 - \nu(a) + \nu(b)) && (14)\\
&\equiv \nu(a \implies_R b) && (15)
\end{aligned}
$$

  which shows contrapositive symmetry. As a result, the MT inference in Equation (12) assigning a truth value $\nu(\neg_R H) = 1 - \alpha$ is valid for this R-Implication.

- Godel and product t-norms: from Table 2 we see that the negation for Godel and product t-norms still entail that $\nu(P(E|H) \geq \alpha) = 0$ if we have $\nu(\text{P2}) = 1$. What remains to be assessed is whether such a result leads to a compatibility between our evaluation that $\nu(\text{P1}) = 1 - \alpha$ and the R-Implication rule for Godel's and products t-norms. In

fact this is not the case since both R-Implications will in fact return $\nu(\text{P1}) = 0$ if one assumes $\nu(H) > 0$ or $\nu(\text{P1}) = 1$ if $\nu(H) = 0$. In either case, this is not compatible with the probability-informed truth value estimate of $1 - \alpha$ that we have suggested for premise P1.

In the end, while the Lukasiewicz R-implication gives an identical result to S-implications, it is not possible to conduct the MT inference argument from Equation (8) with the Godel and product t-norms without running into consistency issues. In fact the induced R-negation for such t-norms drastically restricts the acceptable valuation models for P1 and P2 (e.g., it is not possible to have $0 < \nu(\text{P2}) < 1$). One may argue that such restrictions defeat the whole point of using fuzzy logics in the first place, but in fact they simply model different facets of the corresponding classical connectives. For example, a classical conditional is automatically false if the consequent is false and it is conceivable to represent a negation as an operation making exactly false all partial truths. Nonetheless, it is worth exploring how the R-Implications behave when used with the S-negation.

3.3. Modus Tollens Under R-Implications and S-Negation in Fuzzy Logic

We now look at fuzzy logics using R-implications alongside the S-negation. Our result on the Lukasiewicz R-implication is unaffected by such a change since its corresponding R-negation was already the S-negation. We will therefore focus on the problematic ones: the Godel and product R-implications. First, we need to recognise a problem that the residua of each of these t-norms has: if the consequent has truth value zero, it is not possible to obtain a non-zero truth value for their corresponding R-implications and therefore impossible to have $\nu(\text{P1})$. To obtain a non-zero truth value for the consequent in P1, we need to relax the absolute truth value of P2. Assuming there is some precision—usually very small compared to the signal itself—in any measurement, we shall posit more generally that $\nu(\text{P2}) = 1 - p$ where p represents the error due to the limited precision. From the S-negation it follows that $\nu(P(E|H) \geq \alpha) = p$ (Note that with the Godel and Product R-negation, the result would have been strictly zero instead). We can now try to see how this might affect our previous conclusions on the use of the Godel and product R-implications in MT.

- Godel t-norm: from Table 2 we see that even if the consequent of the R-implication has non-zero truth value, the truth value of the implication itself can only be either 1 or p in our case. Given that p is a priori independent of α we reach again an inconsistency. For the Godel t-norm, using an S-negation and relaxing the certainty on P2 is not sufficient.

- Product t-norm: from Table 2 we see that if the consequent has valuation p the possible valuations of the implication itself are either 1 if $\nu(H) \leq p$ or $p/\nu(H)$ otherwise. Of these two possible cases, only the second one is compatible with admitting a valuation $1 - \alpha$ for P1. Requiring consistency compels us to get

$$\nu(H) = \min\left(1, \frac{p}{1 - \alpha}\right) \tag{16}$$

where the min function ensures that $\nu(H) \leq 1$. Having used consistency to find a truth value to H we can check whether or not it is consistent with the valuation proposed in Equation (8). To do so, we need first to evaluate the truth value of the contrapositive of a generic R-implication:

$$\nu(\neg_S b \Longrightarrow_R \neg_S a) \equiv \min\left(1, \frac{1 - \nu(a)}{\nu(\neg_S b)}\right) = \min\left(1, \frac{1 - \frac{1 - \nu(\neg_S b)}{\nu(a \Longrightarrow_R b)}}{\nu(\neg_S b)}\right) \tag{17}$$

where the last equality is obtained from the product R-implication under the condition that $\nu(a \implies b) < 1$. We can then substitute into the compositional extension rule for MT:

$$
\begin{aligned}
\nu(\neg_S a) &= \top(\nu(\neg_S b \implies \neg_S a), \nu(\neg_S b)) = \nu(\neg_S b) \times \min\left(1, \frac{1 - \frac{1 - \nu(\neg_S b)}{\nu(a \implies_R b)}}{\nu(\neg_S b)}\right) \\
&= \min\left(\nu(\neg_S b), 1 - \frac{1 - \nu(\neg_S b)}{\nu(a \implies_R b)}\right) \\
&= 1 - \frac{1 - \nu(\neg_S b)}{\nu(a \implies_R b)}
\end{aligned}
\tag{18}
$$

where the last equality follows from the requirement that $\nu(a \implies_R b) < 1$ and the S-negation. Substituting $\nu(\neg_S b)$ for $1 - p$ and $\nu(a \implies_R b)$ for $1 - \alpha$ in Equation (18) gives us the same result as Equation (16) under S-negation.

Upon using a logic mixing R-implications and S-negation we were then able to lift some issues reported for the product t-norm with R-implications and R-negations. It is worth noting that these changes were not enough to make the MT reasoning with valuation from Equation (8) compatible with Godel's t-norm.

4. Discussion and Conclusions

In this work we have reminded and illustrated the fact that SHT, as it is typically practiced today, is logically flawed. On the one hand Bayesian inference would require two unknown parameters in order to deliver any judgement while, on the other hand, the MT logical skeleton mixed with probabilistic vocabulary makes it an unsound argument within classical logic. To look for a logic on which to ground the intuition behind P value based hypothesis rejection, we have explored t-norm based fuzzy logics under the S- and R-conventions for three of the most commonly employed t-norms. We have found that: (a) a fuzzy extension of MT with the S-implication and S-negation constitute a viable logical framework on which to ground frequentist SHT for all t-norms; (b) when R-implications and R-negations are used, only the Lukasiewicz t-norm can be used for MT while the Godel and product t-norms yield results incompatible with the suggested model; and (c) upon using the S-negation with R-implications and relaxing some assumptions of the model, it was possible to salvage the product t-norm as a possible framework for SHT. These findings suggest that fuzzy logic may be suitable to ground SHT and reflect deeper on the kind of assumptions (including logical axioms, t-norms etc...) necessary for such reasonings to be valid and sound. For example, one may wonder what is so special about fuzzy logic that it appears to bypass the need for the two supplementary parameters that a Bayesian inference would require (see Figure 1). Does this mean that there is *free lunch*? Our current understanding of the fuzzy reasoning which have shown promise (i.e., S-implications and S-negation) suggests that this *free lunch* has two origins: the existence of a robust contrapositive symmetry (derived in Equation (11)) on the one hand and the replacement of statistical probability (about ensembles) by the truth value judgement about an event which has actually occurred (e.g., for stating that $\nu(P2) = 1$) on another hand. Such properties do not exist in statistical reasonings and it is not necessarily desirable that they do; after all, although probabilities and truth values take values between 0 and 1, they are ultimately different things.

If one chooses $\alpha = 0.05$ and has no doubt about the fact that the P value is lower than 0.05, then Equation (12) says that it can be (fuzzy) logically concluded that the hypothesis is false with truth value 0.95. This supports current practice in medical research. It is essential to note that 0.95 is not a probability and, more importantly, that this result strongly depends on the valuations we proposed for P1 and P2. While we do believe that we have provided a sensible rationale for these truth values, fuzzy logic alone does not impose a direction on what they should be. A more radical modelling approach for $\nu(P1)$ could have been

to remain as close as possible to the classical strictly false value for P1 by allowing a deviation from zero which depends only slightly on the value of α. For example, any valuation of the form $\nu_n(\text{P1}) \equiv 1 - \alpha^n$ with $n > 0$ would be perfectly legitimate as far as limiting cases are concerned but would give rise to a drastically different truth value for the conclusion. In that regard, far from being final, we believe the present work opens up research avenues in the application of (fuzzy) logic to the justification and critical appraisal of scientific practice: one, as mentioned before, in the modelling of valuation functions for the premises in the MT inference rule and the other related to the fundamental problem that it is not clear what constitutes a "small enough" value for α to warrant actual decision making. If one can reject a hypothesis with truth value 0.95, to what extent would it be legitimate to reject it if the truth value was instead 0.9? For instance, particle physics works with P values of the order of 3×10^{-7} (one-tailed test) with the 5 sigma convention, which has attracted attention in relation to the announcement of the Higgs boson detection in 2012. These discrepancies show that the whole P value reasoning is based on a somewhat vague concept of "small enough": "if the P value is "small enough" then we can reject the hypothesis". Incidentally, fuzzy logic offers tools which we have not had the opportunity to use in the present paper but which would nonetheless enable one to deal with such vagueness. These are the so-called *linguistic variables* [28] which replace the strict propositions to be evaluated (e.g., "P value less than 0.05") by vague sentences modelled as fuzzy sets (e.g., "P value is small").

In the end, we have showed that, as a proof of principle, it was possible to ground SHT in fuzzy logic and model current practices of frequentist statistics. Far from being the final word on the matter, we hope that this article will spur further works along these lines to delineate the kind of reasonings being used in scientific discourses.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Popper, K. (Ed.) *The Logic of Scientific Discovery*; Routledge: London, UK, 2005.
2. Chalmers, A. *What Is This Thing Called Science?* 3rd ed.; Open University Press: London, UK, 1999.
3. Duhem, P. *The Aim and Structure of Physical Theory*; Princeton University Press: Princeton, NJ, USA, 1954.
4. Lehmann, E.L. The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory of Two? *Am. Stat. Assoc.* **1993**, *88*, 201–208. [CrossRef]
5. Tendeiro, J.N.; Kiers, H.A.L. A review of issues about null hypothesis Bayesian testing. *Psychol. Methods* **2019**, *24*, 274–795. [CrossRef]
6. Stern, J.M.; Izbicki, R.; Esteves, L. G. Stern, R.B. Logically-consistent hypothesis testing and the hexagon of oppositions. *Log. J. IGPL* **2017**, *25*, 241–757. [CrossRef]
7. Stern, J.M.; De Braganca Pereira, C.A. Bayesian epistemic values: focus on surprise, measure probability! *Log. J. IGPL* **2014**, *22*, 236–254. [CrossRef]
8. Lu, C. Channels? Confirmation and Predictions? Confirmation: From the Medical Test to the Raven Paradox. *Entropy* **2020**, *22*, 384. [CrossRef] [PubMed]
9. Fisher, R. Statistical Methods and Scientific Induction. *J. R. Stat. Soc. Ser. B (Methodol.)* **1955**, *17*, 69–78. [CrossRef]
10. Neyman, J.; Pearson, E. On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part I. *Biometrika* **1928**, *20A*, 175–240.
11. Perezgonzalez, J.D. Fisher, Neyman-Pearson or NHST? A tutorial for teaching data testing. *Front. Psychol.* **2015**, *6*, 223. [CrossRef]
12. Greco, D. Significance Testing in Theory and Practice. *Brit. J. Phil. Sci.* **2011**, *62*, 607–637. [CrossRef]
13. Nickerson, R. Null hypothesis significance testing: A review of an old and continuing controversy. *Psychol. Methods* **2000**, *5*, 241–301. [CrossRef] [PubMed]
14. Gigerenzer, G. Mindless statistics. *J. Socio-Econ.* **2004**, *33*, 597–606. [CrossRef]
15. Goodman, S.N. Towards Evidence-Based Medical Statistics. 1: The P Value Fallacy. *Ann. Intern. Med.* **1999**, *130*, 995–1004. [CrossRef]

16.　Consonni, G.; Fouskakis, D.; Liseo, B.; Ntzoufras, I. Prior Distributions for Objective Bayesian Analysis. *Bayesian Anal.* **2018**, *13*, 227–279. [CrossRef]

17.　Sober, E. Testability. *Proc. Addresses Am. Philos. Assoc.* **1999**, *73*, 47–76. [CrossRef]

18.　Neyman, J.; Pearson, E. IX. On the problem of the most efficient tests of statistical hypotheses. *Phi. Trans. R. Soc. Lond. A* **1933**, *231*, 289–337.

19.　Smets, P.; Magrez, P. Implication in fuzzy logic. *Int. J. Approx. Reason.* **1987**, *1*, 327–347. [CrossRef]

20.　Priest, G. The logic of Catuskoti. *Comp. Philos.* **2010**, *1*, 24–54. [CrossRef]

21.　Esteva, F.; Godo, L.; Hajek, P.; Navara, M. Residuated fuzzy logic with an involutive negation. *Arch. Math. Log.* **2000**, *39*, 103–124. [CrossRef]

22.　Zadeh, L.A. Fuzzy Sets. *Inf. Control* **1965**, *8*, 338–353. [CrossRef]

23.　Dubois, D.; Esteva, F.; Godo, L.; Prade, H. Fuzzy-set based logics ? An history-oriented presentation of their main developments. In *Handbook of the Hitsory of Logic 8*; Gabbay, D.M., Woods, J., Eds.; Elsevier: Amsterdam, The Netherlands, 2007.

24.　Gottwald, S. (Ed.) *A Treatise on Many-Valued Logics*; Research Studies Press: Baldock, UK, 2001.

25.　Drieschner, M. (Ed.) *Carl Friedrich von Weizsacker: Major Texts in Physics*; Springer: Berlin/Heidelberg, Germany, 2014.

26.　Nguyen, H.T.; Mukaidono, M.; Kreinovich, V. Probability of implication, logical version of Bayes theorem, and fuzzy logic operations. In Proceedings of the 2002 IEEE World Congress on Computational Intelligence, 2002 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE'02, Proceedings (Cat. No.02CH37291), Honolulu, HI, USA, 12–17 May 2002; Volume 1, pp. 530–535. [CrossRef]

27.　Magrez, P.; Smets, P. Fuzzy Modus Ponens: A New Model Suitable in Knowledge Based Systems. *Int. J. Intell. Syst.* **1989**, *4*, 181–200. [CrossRef]

28.　Zadeh, L.A. Is There a Need for Fuzzy Logic? *Inf. Sci.* **2008**, *178*, 2751–2779. [CrossRef]