Acta Analytica
https://doi.org/10.1007/s12136-024-00597-0

Truth-Ratios, Evidential Fit, and Deferring to Informants with Low Error Probabilities

Michael Roche[1] · **William Roche**[2]

Received: 19 December 2023 / Accepted: 20 May 2024

Abstract

Suppose that an informant (test, expert, device, perceptual system, etc.) is unlikely to err when pronouncing on a particular subject matter. When this is so, it might be tempting to defer to that informant when forming beliefs about that subject matter. How is such an inferential process expected to fare in terms of *truth* (leading to true beliefs) and *evidential fit* (leading to beliefs that fit one's total evidence)? Using a medical diagnostic test as an example, we set out a formal framework to investigate this question. We establish seven results and make one conjecture. The first four results show that when the test's error probabilities are low, the process of deferring to the test can score well in terms of (i) both truth and evidential fit, (ii) truth but not evidential fit, (iii) evidential fit but not truth, or (iv) neither truth nor evidential fit. Anything is possible. The remaining results and conjecture generalize these results in certain ways. These results are interesting in themselves—especially given that the diagnostic test is not sensitive to the target disease's base rate—but also have broader implications for the more general process of deferring to an informant. Additionally, our framework and diagnostic example can be used to create test cases for various reliabilist theories of inferential justification. We show, for example, that they can be used to motivate evidentialist process reliabilism over process reliabilism.

Keywords Base-rate fallacy · Bayes's theorem · Error probabilities · Evidential fit · Deference to informants · Inferential justification · Reliabilism · Reliability · Truth

✉ Michael Roche
rochmich@isu.edu

William Roche
w.roche@tcu.edu

1 Department of English and Philosophy, Idaho State University, Pocatello, ID, USA

2 Department of Philosophy, Texas Christian University, Fort Worth, TX, USA

Published online: 20 June 2024

1 Introduction

Consider a doctor ("Doc") who follows test T when determining whether his patients have disease D. If T says of a given patient S that she has D, then Doc infers that she has D. If T says of a given patient S that she does not have D, then Doc infers that she does not have D. Call this process "Follow-T". Suppose that T's error probabilities are low in that:

$$\Pr(\text{T says that S does not have D} \mid \text{S has D}) = 0.05 \quad (1.1)$$

$$\Pr(\text{T says that S has D} \mid \text{S does not have D}) = 0.05 \quad (1.2)$$

These probabilities entail that T is both "highly sensitive"—rarely leading to false negatives—and "highly specific"—rarely leading to false positives. However, do these probabilities make it probable that Follow-T will have a high *truth-ratio* (i.e., a high ratio of true beliefs produced to total beliefs produced)? And, given those probabilities, how will Follow-T fare in terms of *evidential fit* (leading to beliefs that fit the total evidence)?

Because Follow-T is an inferential (or "belief-dependent") process, the truth-ratios at issue are *conditional* truth-ratios. A process's conditional truth-ratio is its truth-ratio restricted to cases where all the input beliefs are true.[1] For ease of presentation, however, we shall suppress this detail in what follows and speak simply of "truth-ratios".

Two virtues of Follow-T are apparent. First, when Doc uses Follow-T, his diagnostic judgments are based on evidence—evidence about what T says. Second, that evidence always discriminates between the hypothesis that S has D and the hypothesis that S does not have D in that:

$$\begin{aligned}&\Pr(\text{T says that S has D} \mid \text{S has D}) = 0.95 > \\ &0.05 = \Pr(\text{T says that S has D} \mid \text{S does not have D})\end{aligned} \quad (1.3)$$

$$\begin{aligned}&\Pr(\text{T says that S does not have D} \mid \text{S has D}) = 0.05 < \\ &0.95 = \Pr(\text{T says that S does not have D} \mid \text{S does not have D})\end{aligned} \quad (1.4)$$

Things would be quite different if Doc used, for example, random guessing or crystal ball gazing. If he used the former, then his diagnostic judgments would not be based on evidence (or anything else). If he used the latter, then though his diagnostic judgments would be based on evidence—evidence about what his crystal ball says—that evidence would be worthless in that it would never discriminate between the hypothesis that S has D and the hypothesis that S does not have D.[2]

Doc's process is not beyond reproach, however. For Follow-T can lead Doc to form beliefs that fail to fit his total evidence. Let "S" be one of Doc's patients, let "Has" be the claim that S has D, and let "Says" be the claim that T says that S has D.

[1] See Goldman (1979) on conditional reliability versus unconditional reliability.

[2] For discussion of discrimination, favoring, and how to measure them, see, for example, Edwards (1992), Hacking (1965), Roche and Sober (2019), Royall (1997) and Sober (2008, 2009).

Suppose that S is a random member of the population from Doc's perspective, and that Doc knows that D's base rate in the population is 0.0001 (and has no additional evidence relevant to the matter). Given all this, Doc's prior probability for Has is (or should be) 0.0001. It follows by Bayes's theorem that Pr(Has | Says) roughly equals 0.002.[3] Thus, if and when Doc learns Says and uses Follow-T to infer Has, Follow-T leads him to form a belief that fails to fit his total evidence, for the probability of Has given his total evidence is equal to Pr(Has | Says) and thus is very close to 0.[4]

This is closely related to the base-rate fallacy. This probabilistic fallacy can be formulated as follows (where "E" is some piece of evidence and "H" is some hypothesis):

$$\Pr(\sim E \mid H) \text{ is low.}$$

$$\Pr(E \mid \sim H) \text{ is low.}$$

Thus

$$\Pr(H \mid E) \text{ is high.}$$

This is a fallacy in that there can be cases where the premises are true and the conclusion is false.[5] The case of Doc is a case in point. Pr(~Says | Has) and Pr(Says | ~Has) are both low—these are the error probabilities from (1.1) and (1.2)—but so too is Pr(Has | Says).[6]

It is far from clear, then, how Follow-T fares in terms of truth and evidential fit. When Doc uses Follow-T (and the error probabilities and base rate are as specified above), he takes into account evidence that discriminates between the hypothesis that S has D and the hypothesis that S does not have D. This is some reason to think that Follow-T is likely to have a high truth-ratio. Yet, when doing so, he fails to take into account prior probabilities and can therefore be led to form beliefs that do not fit his total evidence. This is some reason to think that Follow-T scores poorly in terms of evidential fit.[7]

3 Pr(Has | Says) = [Pr(Has)Pr(Says | Has)] / Pr(Says) = [(0.0001)(0.95)] / 0.05009 ≈ 0.002. Here, Pr(Says) = Pr(Has)Pr(Says | Has) + Pr(~Has)Pr(Says | ~Has).

4 We assume that a belief that H fits one's total evidence if and only if the probability of H given one's total evidence is high. Note, however, that we do not assume that if a belief fails to fit one's total evidence, then that belief is unjustified. That is a separate issue. We shall focus on justification in Section 4.2. See Feldman and Conee (1985) for defense of a theory of justification on which a certain notion of "fittingness" plays a central role.

5 Our characterization of the base-rate fallacy is similar to the characterizations in Howson and Urbach (2006, p. 24) and Psillos (2007, pp 17–18).

6 We are assuming that 0.05 is a low value. But there is nothing essential in this. For any value v such that $0.05 > v > 0$, there are cases where (a) Pr(~Says | Has) and Pr(Says | ~Has) are both equal to v and (b) Pr(Has | Says) is less than or equal to 0.5 and thus is not high. For example, when Pr(~Says | Has) and Pr(Says | ~Has) are both equal to 0.0005 and Pr(Has) is equal to 0.0000001, Pr(Has | Says) is roughly equal to 0.0002.

7 Isaacs (2021) makes a similar point about "calibrationism" (or "calibrationism schema"), where this is the view that "[i]f the expected reliability of an agent's judgment regarding p is r, then if that agent judges that p the agent's credence in p should be r" (p. 248). He argues in particular that calibrationism relies on the base-rate fallacy.

Our interest in Follow-T is due to our interest in the following belief-forming strategy, of which Follow-T is an instance:

> Follow-Informant: For any proposition P in domain M, if informant INF says that P is true, then believe that P is true.

This should be understood broadly so that INF can be a person (e.g., a physicist), a test (e.g., T), a device (e.g., a fuel gauge), and even a perceptual system (e.g., a visual system). What should we make of Follow-Informant? If you want to have belief-forming processes that score well in terms of truth and/or evidential fit, should you use Follow-Informant? We shall see, in Section 4.1, that the answer to this question is "not necessarily", for even when INF's error probabilities are low, it need not be the case that Follow-Informant will score well in terms of truth or evidential fit. Since Follow-T is an instance of Follow-Informant, the same can be said of it.

The issues here are not of mere theoretical interest. For it is not uncommon to find ourselves relying on the pronouncements of informants. Although we shall focus mostly on Follow-T in the next two sections, bear in mind that what we say about it bears more generally on Follow-Informant and other instances thereof.

Follow-Informant is not the only way to make use of an informant's pronouncements. In Section 4.1, we describe a Bayesian strategy, which we return to briefly at the end of Section 4.2. The latter is not our focus in what follows, but not because we take it to be inferior to or less interesting than Follow-Informant. Quite the opposite. We focus on Follow-Informant in part due to the fact that it has been underexplored in comparison to Bayesian strategies.

The question of how Follow-T scores, or is likely to score, in terms of truth and evidential fit is underspecified at this point. We shall remedy this by setting out, in Section 2, a framework called "Constant" and reformulating the question in terms of it. We shall then show, in Section 3, that the situation is mixed. There are instances of Constant in which Follow-T scores well in terms of both truth and evidential fit. But there are also instances in which Follow-T scores well in terms of truth but not evidential fit, instances in which it scores well in terms of evidential fit but not truth, and instances in which it does not score well in terms of either evidential fit or truth. We shall also show that some of these results can be generalized in various respects by relaxing certain constraints in Constant.

Though these results are interesting in themselves, they are also interesting because of how they can be brought to bear on various issues in epistemology and philosophy of science. In Section 4.1, we show how these results bear on the viability of Follow-Informant. In Section 4.2, we show how these results can be used to construct test cases for various reliabilist theories of justification. In particular, we show that some such cases suggest that "process reliabilism" is inferior to "evidentialist process reliabilism", and that others suggest that the latter is inferior to "indicator reliabilism".

2 Questions

We mentioned above the need to introduce a framework for making precise our target question. That framework is as follows:

> Constant: (i) Doc will use T on each member of a random sample of (finite) size n of a given population; (ii) Pr(Has) is *constant* across all uses of T; (iii) Pr(~Says | Has) and Pr(Says | ~Has) are *constant* across all uses of T; and (iv) Pr(Has), Pr(~Says | Has), and Pr(Says | ~Has) are greater than 0 and less than 1.

This is a natural place to start, given its simplicity. In reality, of course, T's error probabilities could vary across time, and thus across uses of T, and, similarly, Pr(Has) could vary from one test subject to another.

Constant is a general framework for generating specific sets of conditions. When values for n, Pr(Has), Pr(~Says | Has), and Pr(Says | ~Has) are specified, the result is a specific set of conditions, and that set is an instance of Constant.

We can now state our four main target questions:

> Question 1: Are there instances of Constant such that T's error probabilities are low, it *is* highly probable that Follow-T will have a high truth-ratio, and Follow-T *cannot* lead Doc to form beliefs that fail to fit his total evidence?

> Question 2: Are there instances of Constant such that T's error probabilities are low, it *is* highly probable that Follow-T will have a high truth-ratio, and Follow-T *can* lead Doc to form beliefs that fail to fit his total evidence?

> Question 3: Are there instances of Constant such that T's error probabilities are low, it is *not* highly probable that Follow-T will have a high truth-ratio, and Follow-T *cannot* lead Doc to form beliefs that fail to fit his total evidence?

> Question 4: Are there instances of Constant such that T's error probabilities are low, it is *not* highly probable that Follow-T will have a high truth-ratio, and Follow-T *can* lead Doc to form beliefs that fail to fit his total evidence?

The first and last of these questions ask about *pure* cases, where Follow-T fares well (or poorly) in terms of both truth and evidential-fit. The other two questions ask about *mixed* cases, where Follow-T fares well in terms of truth but not evidential-fit or vice versa.

We shall assume, for simplicity, that the threshold for a high probability equals the threshold for a high truth-ratio. We call this "**t**" and stipulate that $0.5 \leq \mathbf{t} < 1$. We shall also assume that this threshold determines the threshold for a low error probability in that the latter equals $1 - \mathbf{t}$.[8] Finally, we shall assume for now, that $\mathbf{t} = 0.95$.

[8] This assumption is fairly natural, for "high probability" is a measure of closeness to 1 and "low probability" is a measure of closeness to 0. If one regards, say, 0.85 to be a reasonable threshold for high probability, it is natural to regard 0.15 to be a reasonable threshold for low probability; after all, each is the same distance from 1 and 0, respectively.

Consequently, a high probability is greater than 0.95, a high truth-ratio is greater than 0.95, and a low error probability is less than 0.05. Note, however, that we consider alternative values for **t** at the end of Section 3.

Our context of interest differs importantly from contexts in which the aim is to try to *estimate* an instrument's error probabilities by repeated usages of it. We are *assuming* specific values for T's error probabilities, and considering some "downstream" questions based on that assumption. We think that estimation contexts are interesting and important, but require separate treatment (see Osimani & Landes, 2023 for recent discussion).

The questions above are framed in terms of *high probabilities of high truth-ratios* as opposed to just *high truth-ratios*. The reason for this can be seen by returning to random guessing. There are clearly cases in which random guessing happens to have a high truth-ratio. After all, a random guess can be true by luck, and so too can all random guesses. However, random guessing nonetheless has no viability when it comes to the goal of truth. The situation would be different, we take it, if it was highly probable in a given case for random guessing to have a high truth-ratio. Hence, though there are clearly instances of Constant in which Follow-T has a high truth-ratio, we want to set aside this fact as rather uninformative with respect to its viability, and focus on whether it is highly probable that Follow-T will have a high truth-ratio. If, say, *all* instances of Constant were such that it is highly probable that Follow-T will have a high truth-ratio, then this would give Follow-T a significant leg up on random guessing.

There is no mention of "reliability" in the four questions above. But for each of those questions, and for each of the three reliabilist theories noted above in Section 1, there is a corresponding reliability question about Follow-T. We shall return to this in Section 4.2.

Our next task is to answer Question 1 - Question 4.

3 Results

We establish seven results in this section. We also make one conjecture. The first four results are answers to Question 1 - Question 4. The remaining results and the conjecture are in some ways more general than those results.

Constant specifies that Doc uses Follow-T on each member of a random sample of n people. This means that he runs T on each member of the sample, notes what T says in each case, and comes to believe whatever it is that T says. He thus uses Follow-T n times to form n beliefs. We can think of each usage as a "trial", where successes are true beliefs and failures are false beliefs. The probability of a success in a given trial is given by:

$$\begin{aligned} p &= \Pr(\text{Has\&Says}) + \Pr(\sim \text{Has\&} \sim \text{Says}) \\ &= \Pr(\text{Has})\Pr(\text{Says} \mid \text{Has}) + \Pr(\sim \text{Has})\Pr(\sim \text{Says} \mid \sim \text{Has}) \\ &= \Pr(\text{Has})\left[1-\Pr(\sim \text{Says} \mid \text{Has})\right] + [1-\Pr(\text{Has})]\left[1-\Pr(\text{Says} \mid \sim \text{Has})\right] \end{aligned} \tag{3.1}$$

Table 1 $Constant_1$, $Constant_2$, $Constant_3$, and $Constant_4$

	$Constant_1$	$Constant_2$	$Constant_3$	$Constant_4$
n	10,000	10,000	10,000	10,000
Pr(HAS)	= 0.5	= 0.001	= 0.5	= 0.001
Pr(~SAYS ǀ HAS)	= 0.025	= 0.025	= 0.04999999	= 0.04999999
Pr(SAYS ǀ ~HAS)	= 0.025	= 0.025	= 0.04999999	= 0.04999999

Constant also specifies that Pr(HAS), Pr(~SAYS ǀ HAS), and Pr(SAYS ǀ ~HAS) are constant across trials. It follows that this is a binomial experiment in which p is fully determined by these three probabilities.[9]

The probability of *exactly x* successes in a binomial experiment is given by:

$$\Pr(=x) = \left({}_{n}C_{x}\right)(p)^{x}(1-p)^{n-x} \tag{3.2}$$

Here "${}_{n}C_{x}$" is the number of ways in which there can be exactly x successes in n trials, where this equals:

$$n!/x!(n-x)! \tag{3.3}$$

Given (3.2), it follows that the probability of *more than x* successes in a binomial experiment is given by:

$$\Pr(>x) = \sum_{i=x+1}^{i=n} \left({}_{n}C_{i}\right)\left(p^{i}\right)(1-p)^{n-i} \tag{3.4}$$

In the case of Follow-T, we are interested in the probability of its having a high (> **t**) truth-ratio. Thus, the probability at issue is given by:

$$\Pr(>\mathbf{t}) = \sum_{i=nt+1}^{i=n} \left({}_{n}C_{i}\right)\left(p^{i}\right)(1-p)^{n-i} \tag{3.5}$$

Since this probability is fully determined by the values for n and p, and since, as noted above, p is fully determined by Pr(HAS), Pr(~SAYS ǀ HAS), and Pr(SAYS ǀ ~HAS), it follows that whether it is highly probable that Follow-T will have a high truth-ratio is fully determined by the values for n, Pr(HAS), Pr(~SAYS ǀ HAS), and Pr(SAYS ǀ ~HAS).

We mentioned at the end of the previous section that we would consider four instances of Constant. These instances are set out in Table 1. Table 2 lays out, for each of these instances of Constant, values for Pr(> **t**), Pr(HAS ǀ SAYS), and Pr(~HAS

[9] Binomial experiments have four essential characteristics. First, there is a fixed number of trials: "n". Second, each trial has the same two possible outcomes: "success" and "failure". Third, the probability of a success in any given trial is probabilistically independent of the outcomes in the other trials. Fourth, the probability of a success is the same for every trial: "p". See any standard text book on statistics for an introduction to binomial experiments.

Table 2 Pr(> t), Pr(Has | Says), and Pr(~Has | ~Says) in $Constant_1$ – $Constant_4$

	$Constant_1$	$Constant_2$	$Constant_3$	$Constant_4$
Pr(> **t**)	≈ 1.000	≈ 1.000	≈ 0.494	≈ 0.494
Pr(Has | Says)	= 0.975	≈ 0.038	≈ 0.95000001	≈ 0.019
Pr(~Has | ~Says)	= 0.975	≈ 0.99997	≈ 0.95000001	≈ 0.99995

[1]Appendix A provides details for calculating the values of Pr(> **t**) across the four cases. The values for Pr(Has | Says) and Pr(~Has | ~Says) can be calculated using Bayes's theorem from footnote 3 from Section 1. This note applies to Table 3 (below) as well.

| ~Says). Recall that we are assuming (for now) that **t** = 0.95. This is the threshold for both a high probability and a high truth-ratio. Recall as well that we assume that this implies that the threshold for a low probability is 1 – **t** = 0.05. Given these assumptions, Follow-T's error probabilities are low in all four instances of Constant. Yet while it is highly probable that Follow-T will have a high truth-ratio in $Constant_1$ and $Constant_2$, it is *not* highly probable that Follow-T will have a high truth-ratio in $Constant_3$ and $Constant_4$. (The first row of Table 2 shows this.) And while Follow-T cannot lead Doc to form beliefs that fail to fit his total evidence in $Constant_1$ and $Constant_3$, Follow-T *can* lead Doc to form beliefs that fail to fit his total evidence in $Constant_2$ and $Constant_4$. (The second and third rows of Table 2 show this.)[10]

This all suffices to show:

> Result 1: There are instances of Constant such that T's error probabilities are low, it *is* highly probable that Follow-T will have a high truth-ratio, and Follow-T *cannot* lead Doc to form beliefs that fail to fit his total evidence.

> Result 2: There are instances of Constant such that T's error probabilities are low, it *is* highly probable that Follow-T will have a high truth-ratio, and Follow-T *can* lead Doc to form beliefs that fail to fit his total evidence.

> Result 3: There are instances of Constant such that T's error probabilities are low, it is *not* highly probable that Follow-T will have a high truth-ratio, and Follow-T *cannot* lead Doc to form beliefs that fail to fit his total evidence.

> Result 4: There are instances of Constant such that T's error probabilities are low, it is *not* highly probable that Follow-T will have a high truth-ratio, and Follow-T *can* lead Doc to form beliefs that fail to fit his total evidence.

Hence the answer to each of our four main target questions is affirmative.

It is crucial to bear in mind that we are using the expressions "highly probable" and "high" in a stipulative sense, not an everyday sense. (See the end of Section 2.) Consider Result 3, for example. It claims that there are instances of Constant such that T's error probabilities are less than 0.05, the probability that Follow-T will have

[10] Recall that we understand fit such that S's belief that H fits her total evidence if and only if the probability of H given her total evidence is high.

a truth-ratio greater than 0.95 is not greater than 0.95, and Follow-T cannot lead Doc to form beliefs that fail to fit his total evidence. This leaves it open that in the instances in question, the probability that Follow-T will have a truth-ratio greater than 0.94, say, *is* greater than 0.95. If this were the case, it would be natural in an everyday setting to say that it is highly probable that Follow-T will have a high truth-ratio. But this would not be problematic for Result 3, properly understood. (We shall consider some alternative values for **t** at the end of this section.)

It might seem odd that there are instances of Constant such that it is highly probable that Follow-T will have a high truth-ratio even though Doc ignores Pr(Has) when he uses it. It turns out, however, that Pr(Has) sometimes plays *no* role in determining whether it is highly probable that Follow-T will have a high truth-ratio. Suppose, as in the four instances of Constant above, that Pr(~Says | Has) and Pr(Says | ~Has) are equal to each other. Let their value be "e". It follows that:

$$\begin{aligned} p &= \Pr(\text{Has\&Says}) + \Pr(\sim \text{Has\&} \sim \text{Says}) \\ &= \Pr(\text{Has})\Pr(\text{Says} \mid \text{Has}) + \Pr(\sim \text{Has})\Pr(\sim \text{Says} \mid \sim \text{Has}) \\ &= \Pr(\text{Has})(1\text{–}e) + \Pr(\sim \text{Has})(1\text{–}e) \\ &= (1\text{–}e)[\Pr(\text{Has}) + \Pr(\sim \text{Has})] \\ &= 1\text{–}e \end{aligned} \tag{3.6}$$

Given this, and given that whether it is highly probable that Follow-T will have a high truth-ratio is fully determined by the values for n and p, it follows that when T's two error probabilities are equal to each other, Pr(Has) plays no role in determining whether it is highly probable that Follow-T will have a high truth-ratio.

This leads to two supplementary results:

Result 5: For all values for Pr(Has), there are instances of Constant such that T's error probabilities are low and it *is* highly probable that Follow-T will have a high truth-ratio.

Result 6: For all values for Pr(Has), there are instances of Constant such that T's error probabilities are low and it is *not* highly probable that Follow-T will have a high truth-ratio.

First, take Constant_1 and modify it in terms of any alternative value for Pr(Has). Given that T's two error probabilities are equal to each other, it follows that Pr(> 0.95) remains unchanged. Hence it *is* highly probable that Follow-T will have a high truth-ratio. This suffices to establish Result 5. Second, take Constant_3 and modify it in terms of any alternative value for Pr(Has). Given that T's two error probabilities are equal to each other, it follows that Pr(> 0.95) remains unchanged. Hence it is *not* highly probable that Follow-T will have a high truth-ratio. This suffices to establish Result 6.

The situation is quite different when it comes to whether Follow-T can lead Doc to form beliefs that fail to fit his total evidence. Consider Constant_1 and Constant_2, for example. Follow-T can lead Doc to form beliefs that fail to fit his total evidence in the latter but not in the former. This is because Pr(Has | Says) is close to 0 in Constant_2, whereas both Pr(Has | Says) and Pr(~Has | ~Says) are close to 1 in Constant_1. This, in turn, is because Pr(Has) equals 0.001 in Constant_2, whereas it equals 0.5 in Constant_1. So, though, in these cases, the value for Pr(Has) plays no role in determining whether it is highly probable that Follow-T will have a high truth-ratio, it plays a significant role in determining whether Follow-T can lead Doc to form beliefs that fail to fit his total evidence.

Result 5 and Result 6 are like the first four results in that they should be understood so that $\mathbf{t} = 0.95$, and thus the threshold for low probabilities equals 0.05. They are also like the first four in that they are established by appeal to instances of Constant in which n equals 10,000. What about other values for **t** and n?

It turns out that Result 5 generalizes to *all* values for **t** and n such that $1 > \mathbf{t} \geq 0.5$ and $\infty > n > 0$:

> Result 7: For all values for Pr(Has), all values for **t** such that $1 > \mathbf{t} \geq 0.5$, and all values for n such that $\infty > n > 0$, there are instances of Constant such that T's error probabilities are low and it *is* highly probable that Follow-T will have a high truth-ratio.[11]

This is proven in Appendix B.

Now what about Result 6? Does it generalize in the way that Result 5 does? We are not sure at this point. But here is a conjecture:

> Conjecture: For all values for Pr(Has), all values for **t** such that $1 > \mathbf{t} \geq 0.5$, and all values for n such that $\infty > n > 0$, there are instances of Constant such that T's error probabilities are low and it is *not* highly probable that Follow-T will have a high truth-ratio.

Consider Table 3. There are in effect nine general cases in it. The value for **t** is very high (0.99) in three, middling (0.75) in three, and very low (0.5) in three—recall the constraint that $0.5 \leq \mathbf{t} < 1$. The value for n is very high (10,000,000) in three, middling (1,000) in three, and very low (100) in three. In each case, Pr(~Says | Has) is equal to Pr(Says | ~Has), and thus the value for Pr(Has) has no impact on the value for Pr(> **t**); this is why that value is left indeterminate. Since in all such cases, T's error probabilities are low given the specified value for **t**, but Pr(> **t**) is not greater than **t** given that specified value, this (along with the fact that we have not found any counterexamples in our searches) is at least suggestive that Conjecture is true.

[11] In fact, this result holds for all values for **t** less than 1 but greater than 0. We formulate Result 7 as we do since **t** is the threshold for a high probability and a high truth-ratio, and since 1 – **t** is the threshold for a low probability. So defined, it would be quite odd to set **t** below 0.5.

Table 3 Test cases for Conjecture

t	n	Pr(HAS)	Pr(~SAYS | HAS)	Pr(SAYS | ~HAS)	Pr(> **t**)
0.99	10,000,000	0 ≤ 1	0.00999999	0.00999999	≈ 0.49997
	1,000				≈ 0.458
	100				≈ 0.366
0.75	10,000,000	0 ≤ 1	0.24999999	0.24999999	≈ 0.4999
	1,000				≈ 0.488
	100				≈ 0.462
0.5	10,000,000	0 ≤ 1	0.49999999	0.49999999	≈ 0.4999
	1,000				≈ 0.488
	100				≈ 0.462

A brief summary is in order. First, it follows from Result 1, Result 2, Result 3, and Result 4 that the answer to each of the target questions set out in Section 2 is affirmative. Second, Result 5 and Result 6 show that the truth-ratio parts of Result 1, Result 2, Result 3, and Result 4 generalize to all values for Pr(HAS). Third, Result 7 shows that Result 5 generalizes to all values for **t** and n (given certain minor constraints). Fourth, Result 6 generalizes in the same way if Conjecture is true. There is some reason to think that it is in fact true, but, as its name indicates, it is just a conjecture at this point.

We have ignored the issue of whether the evidential-fit parts of Result 1, Result 2, Result 3, and Result 4 generalize to all values for Pr(HAS), **t**, and n. But this is worth exploring in the future.

4 Discussion

We set out several results in the previous section concerning Follow-T, truth-ratios, and evidential fit. Those results are interesting in themselves. But they are also interesting in terms of how they bear on the viability of Follow-Informant and reliabilist theories of justification. In this section, we discuss these matters in turn.

4.1 Follow Informant

Recall Follow-Informant (from Section 1):

> Follow-Informant: For any proposition P in domain M, if informant INF says that P is true, then believe that P is true.

Because Follow-T is an instance of Follow-Informant, the results from Section 3 bear on the viability of Follow-Informant.

Result 3 shows that there are instances of Follow Informant such that INF's error probabilities are low, but it is not highly probable that that instance will have a high truth-ratio. Result 2 and Result 4 show that there are instances of Follow Informant such that INF's error probabilities are low, but that instance can lead a subject to form beliefs that fail to fit her total evidence. Consequently, if you want a high probability of attaining a high truth-ratio, or if you want to avoid beliefs that do not fit your total evidence, then whether you should use Follow-Informant depends on more than just whether INF's error probabilities are low. Whether you should want a high probability of attaining a high truth-ratio, or whether you should want to avoid beliefs that do not fit your total evidence, is not something we shall weigh in on here.

Let Follow-Informant's "fit-ratio" be the number of beliefs that it outputs that fit the total evidence divided by the number of beliefs that it outputs in total. When Follow-Informant can lead you to form beliefs that fail to fit your total evidence, its fit-ratio can be less than 1. This leaves it open, though, that its fit-ratio is high. It also leaves it open that it is highly probable that its fit-ratio is high. Do low error probabilities for INF guarantee that it is highly probable that Follow-Informant's fit-ratio is high?

The answer is no. Let us stipulate that the threshold for high probability and high fit-ratio is 0.95. And let a "success" in a given trial be a belief that fits the total evidence. Consider $\text{Constant}_{4.}$ Here, Pr(Has | Says) is roughly equal to 0.019 and Pr(~Has | ~Says) is roughly equal to 0.99995. This implies that there is a success in a given trial precisely when ~Says is true. Hence the probability of a success in a given trial is equal to the probability of ~Says. This probability is slightly less than 0.9492. But when the probability of a success in a given trial in a binomial experiment with 10,000 trials equals 0.9492, the probability of more than 9,500 successes is roughly equal to 0.352 and thus fails to be high. Hence, though T has low error probabilities, it is not highly probable that Follow-T's fit-ratio is high. Regardless, then, of whether INF's error probabilities are low, if you care about high fit-ratios, it could be that you should look for an alternative process.

It is worth emphasizing that Follow-Informant is *not* the only way for you to make use of what INF says. If INF says P, then you could use Bayes's theorem along with P's prior probability and INF's error probabilities to determine Pr(P | INF says P), and infer P if and only if this probability is sufficiently high.[12] This Bayesian strategy requires that you know how to use Bayes's theorem, and that you have values for Pr(P) and INF's error probabilities. But it has the advantage that it is guaranteed to have a fit-ratio of 1.

The moral of this section is worth emphasizing. If you care about high truth-ratios and/or high fit-ratios, the fact that INF's error probabilities are low is *insufficient* to establish that you ought to use Follow-Informant, for (i) even when INF's error probabilities are low, it need not be highly probable that Follow-Informant will have a high truth-ratio; and (ii) even when INF's error probabilities are low, it need not be the case that Follow-Informant will have a high fit-ratio.

[12] This should be understood such that "P" stands for any pronouncement from INF. Thus, if INF is T, the above conditional covers both positive (Says) and negative (~Says) cases.

4.2 Reliabilist Theories of Inferential Justification

Because some theories of inferential justification are reliabilist, and because reliability is often understood in terms of truth-ratios, the foregoing discussion can be brought to bear on debates about reliabilist theories of inferential justification.

There are many different reliabilist theories of justification in logical space. Here is a simple version of process reliabilism restricted to inferential justification (since Follow-T is an inferential process):

> Process Reliabilism: If S infers H via process P, then S's belief in H is justified if and only if (i) P is reliable and (ii) all the input beliefs to P are justified.[13]

There are different ways of understanding the reliability condition here. We shall suppose, for definiteness, that the reliability at issue is reliability *in the world in which S infers H*, where P is reliable in that world if and only if *P has a high truth-ratio in that world and nearby possible worlds*. If, for example, S infers H via P in some world W, then condition (i) is satisfied if and only if P has a high truth-ratio in W and nearby possible worlds. This understanding of the reliability condition is fairly standard, though, again, there are others.[14]

For any given use, Follow-T has just one input belief, either a belief in Says or a belief in ~Says. We shall assume, for ease of presentation (and because this is allowed by all the theories discussed in this section), that whenever Doc uses Follow-T, the input belief is justified.

The four instances of Constant set out in Section 3 are not specified in enough detail so that Process Reliabilism issues verdicts on whether Doc's beliefs, formed via Follow-T, are justified. Each of them, though, has a "cousin" case. Constant_1's cousin case, for example, is:

Constant*_1:

(a) Follow-T is used 10,000 times in world W and nearby possible worlds, where each usage is such that Pr(Has) = 0.5, Pr(~Says | Has) = 0.025, and Pr(Says | ~Has) = 0.025.
(b) Pr(Has | Says) = 0.975 and Pr(~Has | ~Says) = 0.975.
(c) Follow-T sometimes leads Doc to infer Has, and sometimes leads him to infer ~Has.
(d) Follow-T's truth-ratio in W and nearby possible worlds is high.[15]
(e) Each of Doc's beliefs formed by Follow-T fits his total evidence.[16]

[13] Goldman (1979, 1986, 2009) defends versions of Process Reliabilism. See also Lyons (2009).

[14] See Comesaña (2009) and Goldman and Beddor (2021) for helpful discussion. See Frise (2018) for complications and challenges.

[15] This is the analogue of the fact in Constant_1 that *it is highly probable* that Follow-T has a high truth-ratio.

[16] This is the analogue of the fact in Constant_1 that Follow-T *cannot* lead Doc to form beliefs that fail to fit his total evidence.

Given (d), Process Reliabilism issues the verdict that all of Doc's beliefs produced by Follow-T are justified. This verdict has some prima facie plausibility, for each belief is formed via a process that has a high truth-ratio in W and nearby possible worlds, and, moreover, each belief fits Doc's total evidence.

But now consider:

Constant*$_2$:

(f) Follow-T is used 10,000 times in W and nearby possible worlds, where each usage is such that Pr(Has) = 0.001, Pr(~Says | Has) = 0.025, and Pr(Says | ~Has) = 0.025.
(g) Pr(Has | Says) ≈ 0.038 and Pr(~Has | ~Says) ≈ 0.99997.
(h) Follow-T sometimes leads Doc to infer Has, and sometimes leads him to infer ~Has.
(i) Follow-T's truth-ratio in W and nearby possible worlds is high.
(j) Some of Doc's beliefs formed by Follow-T, namely, his beliefs in Has, fail to fit his total evidence.

Given (i), Process Reliabilism issues the verdict that all of Doc's beliefs produced by Follow-T are justified. This is prima facie *very* implausible when it comes to his beliefs in Has, for Has's probability given his total evidence is very close to 0.

There is an obvious fix, however. Consider the following evidentialist variant of Process Reliabilism:

> Evidentialist Process Reliabilism: If S infers H via process P, then S's belief in H is justified if and only if (i) P is reliable, (ii) all the input beliefs to P are justified, and (iii) H is highly probable given S's total evidence.[17]

Because Follow-T is reliable in W and Doc's beliefs in ~Has fit his total evidence but his beliefs in Has do not, Evidentialist Process Reliabilism issues the verdict that Doc's beliefs in ~Has are justified but his beliefs in Has are not. Evidentialist Process Reliabilism is thus more discerning than Process Reliabilism. Constant*$_2$ can thus be used to motivate the former view over the latter.

Indeed, such motivation can be more straightforward than alternative motivations. Consider, for example, Goldman's (2011) attempt to motivate a view similar to Evidentialist Process Reliabilism over "single-component" views such as Process Reliabilism and purely evidentialist theories. His argument begins with a case (pp. 263-264): Shirley and Madeleine assign the same degree of belief to a proposition, H. But whereas Shirley does so via a random guess, Madeleine does so by careful and accurate calculation. He then argues as follows:

> Now, on one dimension of justifiedness—the fittingness dimension—Shirley and Madeleine's doxastic attitudes vis-à-vis H deserve the same rating. Equally clearly, however, there is another dimension of justifiedness—call it

[17] Comesaña (2009, 2020) and Goldman (2011) defend versions of Evidentialist Process Reliabilism.

> the process dimension—on which their doxastic attitudes merit different ratings. Madeleine's degree of belief is much more aptly, or competently, chosen than Shirley's—despite the fact that they arrive at the same result. On this second dimension of justifiedness, Shirley's degree is not at all justified or well-founded, whereas Madeleine's degree of belief is very well-founded. A two-factor theory handles this case nicely. But no single-factor theory, of either the purely evidentialist or purely reliabilist sort, can do so. That's a good reason to promote a synthesis of the two. (Goldman, 2011, p. 264)

Goldman's case concerns *degrees of belief* as opposed to *binary beliefs*. But it can be modified in terms of binary beliefs. When so modified, does it favor Evidentialist Process Reliabilism over Process Reliabilism?

No. These theories issue the same verdict: Madeleine's belief is justified but Shirley's belief is not. It is clear, however, from an earlier work (see Goldman, 2009, p. 249) that Goldman means for his "two-component" theory to be understood so that if a belief meets the fittingness condition but not the reliability condition, then it is partly justified but not fully justified, and that if a belief meets both conditions, then it is fully justified and not just partly justified. His theory thus issues the verdict that whereas Madeleine's belief is fully justified, Shirley's belief is partly but not fully justified.

We want to remain neutral on the merits of Goldman's appeal to partial versus full justification. What we wish to point out is that he could bypass this whole issue and simply use a case like Constant*$_2$ to motivate views like Evidentialist Process Reliabilism over views like Process Reliabilism.

Constant*$_2$, it should be noted, is similar in some respects to BonJour's case of Norman the Clairvoyant (1980, pp. 62-65) and Lehrer's case of Mr. TrueTemp (1990, pp. 163-164). Each of the three cases is a case in which a belief is formed by a reliable process but that belief fails to fit the subject's total evidence. There is an important dissimilarity, however. Norman's process and Truetemp's process are perception-like and thus *non-inferential*. Neither of them infers his beliefs from evidence. Follow-T, in contrast, is inferential. Doc forms his beliefs by reasoning from his beliefs about T's pronouncements. Constant*$_2$, but not BonJour's case or Lehrer's case, thus shows how reasoning that is conducive to attaining a high truth-ratio can come apart from reasoning that leads to beliefs that fit the total evidence.

Now consider a simple version of indicator reliabilism:

> Indicator Reliabilism: If S infers H via process P from his belief in E, then S's belief in H is justified if and only if (i) his belief in E is justified and (ii) E is a reliable indicator of H in that Pr(H | E) is high.[18, 19]

---

[18] Alston (1988) and Swain (1981) defend versions of Indicator Reliabilism. See Sturgeon (2000) for discussion.

[19] Indicator Reliabilism should be understood so that Pr(H | E) is "externalist" in that it can be high for one subject in one context and not high for another subject in another context even though the two subjects are identical mentally. Indicator Reliabilism is thus not a mentalist theory as understood by Feldman and Conee (2001).

Since Pr(Has | Says) is low in Constant*$_2$ whereas Pr(~Has | ~Says) is high, Indicator Reliabilism agrees with Evidentialist Process Reliabilism that whereas Doc's beliefs in ~Has are justified, his beliefs in Has are unjustified.

Are there cases in which Indicator Reliabilism disagrees with Evidentialist Process Reliabilism? Yes. Consider:

Constant*$_3$:

(k) Follow-T is used 10,000 times in W and nearby possible worlds, where each usage is such that Pr(Has) = 0.5, Pr(~Says | Has) = 0.04999999, and Pr(Says | ~Has) = 0.04999999.
(l) Pr(Has | Says) ≈ 0.95000001 and Pr(~Has | ~Says) ≈ 0.95000001.
(m) Follow-T sometimes leads Doc to infer Has, and sometimes leads him to infer ~Has.
(n) Follow-T's truth-ratio in W and nearby possible worlds is not high.
(o) Each of Doc's beliefs formed by Follow-T fits his total evidence.

Given (n), Evidentialist Process Reliabilism implies that *all* of Doc's beliefs produced by Follow-T are *unjustified* (because the reliability condition is not met), even those that are highly probable given Doc's total evidence. But given (l), Indicator Reliabilism implies that *all* of Doc's beliefs produced by Follow-T are *justified*.

We will not try to adjudicate this disagreement between Evidentialist Process Reliabilism and Indicator Reliabilism. This would take us too far afield. Our point is merely to show how the foregoing discussion can be brought to bear on debates about reliabilist theories of justification.

Recall, though, that we noted at the end of Section 4.1 that Follow-Informant is not the only way for you to make use of what INF says. You could use INF *and* Bayes's theorem: if INF says P, believe that P if and only if Pr(P | INF says P) is sufficiently high. If you followed this Bayesian strategy in Constant*$_3$, then you would infer exactly what Doc infers via Follow-T.[20] Evidentialist Process Reliabilism implies that all such beliefs would be unjustified whereas Indicator Reliabilism implies that all such beliefs would be justified. Those sympathetic to the Bayesian strategy could perhaps use this fact to argue in favor of Indicator Reliabilism over Evidentialist Process Reliabilism.

5 Conclusion

When T has low error probabilities, how well is Follow-T likely to fare in terms of truth (leading to true beliefs) and evidential fit (leading to beliefs that fit the total evidence)? We have argued that the situation is mixed (see Section 3). There are conditions in which T has low error probabilities, Follow-T has a high probability of having

[20] This is because the probabilities in (l) are just the probabilities that Bayes's theorem would output given Pr(Has), Pr(~Says | Has), and Pr(Says | ~Has). Since these probabilities are sufficiently high, you would infer Has whenever Says is true, and you would infer ~Has whenever ~Says is true.

a high truth-ratio, and Follow-T cannot lead Doc to form beliefs that fail to fit his total evidence. This is the good news for Follow-T. But there is bad news as well. For there are also conditions in which T has low error probabilities, but Follow-T does not have a high probability of having a high truth-ratio or can lead Doc to form beliefs that fail to fit his total evidence. This calls into question the viability of Follow-T and other instances of Follow-Informant, even when the informant in question has low error probabilities (see Section 4.1). This, in turn, is potentially problematic for certain reliabilist theories of justification (see Section 4.2).

We want to close by briefly noting three issues for future investigation. The first concerns the extent to which the results in Section 3 can be generalized. Consider Result 7, for example. Can this result be generalized to *all low values for T's error probabilities*? The second concerns the fact that the results in Section 3 are *existential*. They say in effect that there exist instances of Constant with such and such features. They do not say, though, how "frequent" such instances are. This is worth exploring. Third, and finally, there is the issue of the extent to which the results in Section 3, and potential generalizations of them, can be extended beyond the applications in Section 4. For example, do they have application in evolutionary biology on the issue of whether "fallacious" belief-forming processes can be adaptive, and in philosophy of mind on debates over theories of introspection?

Appendix A

We used *Mathematica* (by Wolfram) to determine the values of Pr(> **t**) in Table 2. When something of the form "Probability[x > v, n**t** ≈ BinomialDistribution[n, p]]" is inputted in *Mathematica*, the output produced is the probability of more than n**t** successes in a binomial distribution in which the number of trials is n and the probability of success on a given trial is p. In Constant_1 and Constant_2, since $n = 10000$, $\mathbf{t} = 0.95$, and the probability of success on a given trial is 1 – 0.025, we determined Pr(> t) by inputting "Probability[x > 9500, x ≈ BinomialDistribution[10000, 1 – 0.025]]". The output was:

0.9999999999999999999999999999999999999999999839491532074

In Constant_3 and Constant_4, since $n = 10000$, $\mathbf{t} = 0.95$, and the probability of success on a given trial is 1 – *0.04999999*, we changed the input to "Probability[x > 9500, x ≈ BinomialDistribution[10000, 1 – *0.04999999*]]". The output was:

0.49359519448

For instances of Constant in which n is not equal to 10,000, **t** is not equal to 0.95, or p is something other than 1 – 0.025 and 1 – 0.04999999, inputs to *Mathematica* can be changed accordingly. (These points apply, *mutatis mutandis*, to the calculations in Table 3 as well.)

Appendix B

Take any values for Pr(HAS), **t**, and n such that $1 > \mathbf{t} \geq 0.5$ and $\infty > n > 0$. Follow-T's truth-ratio equals 1 precisely when x = n (where x, recall, is the number of successes). Hence the probability that Follow-T's truth-ratio equals 1 is given by:

$$\Pr(= \mathrm{n}) = ({}_{\mathrm{n}}\mathrm{C}_{\mathrm{n}})(\mathrm{p})^{\mathrm{n}}(1-\mathrm{p})^{\mathrm{n}-\mathrm{n}} \tag{B1}$$

There is only one way for Follow-T to have exactly n successes (in n trials). So:

$$_{\mathrm{n}}\mathrm{C}_{\mathrm{n}} = 1 \tag{B2}$$

Given that Pr(SAYS | ~ HAS) > 0 and Pr(~SAYS | HAS) > 0, p is less than 1. Hence 1 – p is positive and so:

$$(1-\mathrm{p})^{\mathrm{n}-\mathrm{n}} = 1 \tag{B3}$$

(B1), (B2), and (B3) together imply:

$$\Pr(= \mathrm{n}) = (\mathrm{p})^{\mathrm{n}} \tag{B4}$$

Now suppose that Pr(SAYS | ~ HAS) and Pr(~SAYS | HAS) are both equal to e. As shown in (3.6), this supposition entails that:

$$\mathrm{p} = 1-\mathrm{e} \tag{B5}$$

(B4) and (B5) imply:

$$\Pr(= \mathrm{n}) = (1-\mathrm{e})^{\mathrm{n}} \tag{B6}$$

The right-hand side of (B6) is less than 1 when e is greater than 0, but it approaches 1 as e approaches 0:

$$\lim_{e \to 0}(1 - e)^{n} = 1 \tag{B7}$$

Hence when Pr(SAYS | ~ HAS) and Pr(~SAYS | HAS) are both positive and equal to e, the probability that Follow-T's truth-ratio equals 1 approaches 1 as e approaches 0. Hence, as **t** is less than 1, there is a value for e such that the probability that Follow-T's truth-ratio will be greater than **t** is greater than **t**. QED

Acknowledgements Thanks to Borut Trpin and an anonymous referee for the journal for helpful comments on prior versions of the paper.

Declarations

Conflict of Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Alston, W. (1988). An internalist externalism. *Synthese, 74*, 265–283.
BonJour, L. (1980). Externalist theories of empirical knowledge. *Midwest Studies in Philosophy, 5*, 53–73.
Comesaña, J. (2009). What lottery problem for reliabilism? *Pacific Philosophical Quarterly, 90*, 1–20.
Comesaña, J. (2020). *Being rational and being right*. Oxford University Press.
Edwards, A. (1992). *Likelihood*. John Hopkins University Press.
Feldman, R., & Conee, E. (1985). Evidentialism. *Philosophical Studies, 48*, 15–34.
Feldman, R., & Conee, E. (2001). Internalism defended. *American Philosophical Quarterly, 38*, 1–18.
Frise, M. (2018). The reliability problem for reliabilism. *Philosophical Studies, 175*, 923–945.
Goldman, A. (1979). What is justified belief? In G. Pappas (Ed.), *Justification and knowledge* (pp. 1–23). Reidel.
Goldman, A. (1986). *Epistemology and cognition*. Harvard University Press.
Goldman, A. (2009). Replies to discussants. *Grazer Philosophische Studien, 79*, 245–288.
Goldman, A. (2011). Toward a synthesis of reliabilism and evidentialism? Or: Evidentialism's troubles, reliabilism's rescue package. In T. Dougherty (Ed.), *Evidentialism and its discontents* (pp. 254–280). Oxford University Press.
Goldman, A., & Beddor, B. (2021). Reliabilist epistemology. In E. Zalta (Ed.), *Stanford Encyclopedia of Philosophy (Summer 2021 ed.)* Url = https://plato.stanford.edu/archives/sum2021/entries/reliabilism/
Hacking, I. (1965). *The logic of statistical inference*. Cambridge University Press.
Howson, C., & Urbach, P. (2006). *Scientific reasoning: The Bayesian approach* (3rd ed.). Open Court.
Isaacs, Y. (2021). The fallacy of calibrationism. *Philosophy and Phenomenological Research*, *102*, 247–260.
Lehrer, K. (1990). *Theory of knowledge*. Routledge.
Lyons, J. (2009). *Perception and basic beliefs: zombies, modules, and the problem of the external world*. Oxford University Press.
Osimani, B., & Landes, J. (2023). Varieties of error and varieties of evidence in scientific inference. *British Journal for the Philosophy of Science, 74*, 117–170.
Psillos, S. (2007). *Philosophy of science. A-Z*. Edinburgh University Press.
Roche, W., & Sober, E. (2019). Discrimination-conduciveness and observation selection effects. *Philosophers' Imprint, 19*(40), 1–26.
Royall, R. (1997). *Statistical evidence: a likelihood paradigm*. Chapman & Hall.
Sober, E. (2008). *Evidence and evolution: the logic behind the science*. Cambridge University Press.
Sober, E. (2009). Absence of evidence and evidence of absence: evidential transitivity in connection with fossils, fishing, fine-tuning, and firing squads. *Philosophical Studies, 143*, 63–90.
Sturgeon, S. (2000). *Matters of mind: consciousness, reason and nature*. Routledge.
Swain, M. (1981). *Reasons and knowledge*. Cornell University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.