

Conditional Degree of Belief

Jan Sprenger*

December 11, 2015

1 Introduction. Conditional Degree of Belief and Probability in Statistical Models.

The normative force of Bayesian inference is based on the existence of a uniquely rational degree of belief in evidence E given a hypothesis H , that is, $p(E|H)$.

To motivate this claim, think of what a Bayesian inference in science typically looks like. There is a competing set of hypotheses that correspond to different values of an unknown parameter, such as the chance $\theta \in [0, 1]$ of a coin to come up heads. Each of these hypotheses specifies the probability of a certain observation E (e.g., two heads in three tosses). This happens by means of a mathematical function, the *probability density* $\rho_H(E)$. For example, in the case of tossing a coin N times, the Binomial distribution describes the probability of observing k heads and $N - k$ tails if it is assumed that the tosses are independent and identically distributed (henceforth, i.i.d.). Assuming that the probability of heads on any particular throw is $H : \theta = 3/5$, this would amount to $\rho_H(E) = \binom{N}{k}(3/5)^k(2/5)^{N-k}$.

Based on these numbers, a Bayesian reasoner is interested in the posterior degree of belief in the hypothesis that θ equals a certain real number, or that θ falls within an interval $[a, b]$.

Agreement on $p(E|H)$ is required for a variety of reasons. First, the Bayesian's primary measure of evidence is the *Bayes factor* (Kass and Raftery, 1995). It quantifies the evidence for a hypothesis H_0 over a competitor H_1 as the ratio of the probabilities $p(E|H_0)/p(E|H_1)$ —or equivalently, as the ratio of posterior to prior odds. When these degrees of belief may rationally vary, it follows that one may not be able to agree on

*Contact information: Tilburg Center for Logic, Ethics and Philosophy of Science (TiLPS), Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Email: j.sprenger@uvt.nl. Webpage: www.laeuferpaar.de

the degree to which E favors H_0 over H_1 , or vice versa. For the application of Bayesian inference in the sciences, that would be a strong setback. On the other hand, plugging in the probability densities $\rho_{H_0}(E)$ and $\rho_{H_1}(E)$ for $p(E|H_0)$ and $p(E|H_1)$ ensures immediate agreement on the Bayes factor (at least for precise hypotheses about the unknown parameter).

Second, agreement on $p(E|H)$ and $p(E|\neg H)$ is required to back up the various convergence to truth theorems for Bayesian inference (e.g., Gaifman and Snir, 1982). These theorems claim, in a nutshell, that different prior opinions will “wash out” as we continue to sample, ensuring agreement on posterior probabilities. This happens via Bayes’ Theorem:

$$\begin{aligned} p(H|E) &= p(H) \frac{p(E|H)}{p(E)} \\ &= \left(1 + \frac{p(\neg H)}{p(H)} \frac{p(E|\neg H)}{p(E|H)} \right)^{-1} \end{aligned}$$

The idea is that in the long run, the ratio of $p(E|\neg H)$ and $p(E|H)$ will dominate the ratio of prior probabilities. Therefore, $p(H|E)$ will converge either to one or to zero, dependent on whether H or $\neg H$ is true. However, if the degrees of belief in E given H (or $\neg H$) may rationally vary, such convergence results are not guaranteed any more. Any value can result in the limit. This undermines the trustworthiness of Bayesian inference and the decision-theoretic paradigm that posterior probabilities should guide our decisions (Savage, 1972).

Third, proposed solutions of the Problem of Old Evidence usually require that the conditional degree of belief $p(E|H)$ equal the corresponding objective probabilities (for extensive treatments, see Earman, 1992; Sprenger, 2015). Otherwise $p(E|H)$ will be equal to one if E is “old evidence”, e.g., evidence that was known before H was formulated. After all, learning H does not modify our belief that the phenomenon E is real ($p(E|H) = p(E) = 1$). In this case, no Bayesian analysis of confirmation by old evidence will get off the ground because by Bayes’ Theorem, $p(H|E) = p(H)p(E|H)/p(E) = p(H)$. See the case study by Howson (1991, 551–552) for predicting the outcome of a coin toss.

Hence, the main question of our paper is

1. What justifies the equality between conditional degrees of belief and probability densities that is required for meaningful Bayesian inference?

$$p(E|H) = \rho_H(E) \quad \text{(The Equality)}$$

Various scholars derive The Equality from a general epistemic norm. They see it as a requirement of rationality that degrees of belief be calibrated with information about the empirical world (e.g., Lewis, 1980; Williamson, 2007, 2010). For instance, if I know that the coin on my desk is fair, I should assign degree of belief $3/8$ to two heads in three i.i.d. tosses. David Lewis (1980) formalized this intuition in his famous *Principal Principle*: if you know that the objective chance of an event E is x , you should also set your degree of belief in E to x .

Do the chances in statistical inference conform to the description of the Principal Principle? Only partially so. A statistical model H describes the probability of various possible observations E by the probability density $\rho_H(E)$. So the chance of E given H is indeed fixed and objective. For instance, if H is the hypothesis that a coin is fair, then the probability of tossing two heads in three i.i.d. tosses is $\rho_H(E) = 3/8$.

This probability is objective, but it is not ontic—it is no physical chance that describes properties of the real world (Rosenthal, 2004). It is not about the objective chance of an event E in a real experiment, but about the objective chance of E *if the statistical model H were true*. In other words, it is part of the *meaning* of H that the probability of E is $3/8$ (Sprenger, 2010). That is, the sentence

If a fair coin is tossed repeatedly, then the chance of two heads in three i.i.d. tosses is $3/8$.

makes no empirical predictions—it has a distinctly analytical flavor. There need not exist any fair coins in the real world for this sentence to be true. The Principal Principle, by contrast, applies to chances in the real world and is silent on chances that presuppose the truth of an idealized statistical model. Without further qualification, the Principal Principle cannot relate probability densities to conditional degrees of belief and justify The Equality. It is part of the task of this paper to provide such a qualification.

Hence, our paper joins the efforts by philosophers of science to clarify the nature of objective probability in scientific reasoning. Hofer (2007) articulated the research program as follows:

the vast majority of scientists using non-subjective probabilities [...] in their research feel little need to spell out what they take objective probabilities to be. [...] [T]o the extent that we intend to use objective probabilities in explanation or predictions, we owe ourselves an account of what it is about the world that makes the imputation and use of certain probabilities correct. (Hofer, 2007, 550)

While the role of physical chance in explanation and prediction is well-explored (Hofer, 2007; Suarez, 2011a,b), the role of mathematical probability densities and their relation to conditional degree of belief is much less investigated. This paper develops an account of conditional degree of belief and its relation to probability densities that squares well with the common practice of statistical inference, and Bayesian inference in particular. In doing so, I would also like to explain why statisticians have never felt the need to develop a substantive, metaphysical account of objective chance—and to argue that such an account is not needed to do meaningful scientific work.

The rest of the paper is structured as follows. Section 2 argues that the ratio analysis of conditional probability cannot answer the above questions and that it fails as an explanation of our practices in statistical reasoning. Section 3 develops the main constructive proposal: conditional degrees of belief should be interpreted in the counterfactual sense already suggested by Frank P. Ramsey. We then elaborate on how this proposal justifies The Equality, that is, the agreement between conditional degrees of belief and probability densities, and we relate Ramsey’s proposal to the problem of chance-credence coordination. Section 4 explores the implications of our proposal for Bayesian inference and discusses some objections whereas Section 5 wraps up our findings.

2 The Ratio Analysis

According to most textbooks on probability theory, statistics and (formal) philosophy, the conditional probability of an event E given H is defined as the ratio of the probability of the conjunction of both events, divided by the probability of H (Jackson, 1991; Earman, 1992; Skyrms, 2000; Howson and Urbach, 2006).

$$p(E|H) := \frac{p(E \wedge H)}{p(H)} \quad \text{(Ratio Analysis)}$$

Usually, this definition is restricted to the case that $p(H) > 0$. Some textbooks take precaution against this case and define the conditional probability $p(E|H)$ as the factor by which the probability of H , $p(H)$, has to be multiplied in order to obtain $p(E \wedge H)$ as a result. Measure-theoretic approaches define conditional probability via the related concept of a conditional expectation. This is a technicality which need not bother us for the moment.

When applied to conditional degree of belief, the ratio analysis offers a particularly simple way out of the problems sketched in the introduction: conditional degree of

belief is reduced to unconditional degree of belief. The rational degree of belief in E given H is just the ratio of the degree of belief in $E \wedge H$, divided by the degree of belief in H . Hence, we need no separate analysis of conditional degree of belief.

We shall now argue that this proposal is a non-starter. For a detailed coverage of this topic, see the seminal article by Alan Hájek (2003): “What Conditional Probability Could Not Be” and the more recent one by Fitelson and Hájek (2016).

The first observation is that it seems implausible to form conditional degrees of belief via the conjunction of both propositions. It is cognitively very demanding to elicit $p(E \wedge H)$ and $p(H)$ and to calculate their ratio. For determining our rational degree of belief that a fair coin yields a particular sequence of heads and tails, it does not matter whether the coin in question is actually fair. Regardless of our degree of belief in that proposition, we all agree that the probability of two heads in three tosses is $3/8$ *if we suppose that the coin is fair*. The Equality offers an easy and direct alternative to Ratio Analysis: we just set our conditional degree of belief in an event equal to the value assigned by the probability density. Indeed, recent psychological evidence suggests that the ratio analysis is a poor description of how people reason with conditional probabilities, pointing out the necessity of finding an alternative (Zhao et al., 2009).

The second observation is that the ratio analysis of conditional probability fails to cover many important cases of statistical inference. Suppose that we reason about the bias of a coin, with H denoting the hypothesis that the coin is fair. Let our hypothesis space be $\mathcal{H} = \{\theta \in [0, 1]\}$. If our probability density over the elements of \mathcal{H} with respect to the Lebesgue-measure is continuous (e.g., uniform or beta-shaped), then each single point in $[0, 1]$, which corresponds to a particular hypothesis about the bias of the coin, will have probability zero. That is, we are virtually certain that no particular real number describes the true bias of the coin. This sounds quite right given the uncountably many ways the coin could be biased, but paired with Ratio Analysis, it leads to strange results: we cannot assign a probability to a particular outcome (e.g., two heads in three i.i.d. tosses) given that the coin has a particular bias. This sounds outrightly wrong.

Finally, we note that in the paradigmatic cases described above, $p(E|H)$ seems to be constrained in an objective way—constraints which are *prima facie* absent from $p(H)$ and $p(E \wedge H)$. This is another mismatch between the ratio analysis and a concept of conditional degree of belief that is rationally constrained by statistical modeling assumptions.

It shall be concluded that Ratio Analysis cannot deliver a convincing analysis of conditional degree of belief and therefore not answer the question why those degrees

of belief should match probability densities—that is, why The Equality holds.

3 The Counterfactual Analysis

Between the lines, the previous sections have already anticipated an alternative analysis of conditional degree of belief. Rather than conforming to the ratio analysis, we could understand the concept in a *counterfactual* way. That is, we could determine our degrees of belief in E given H by supposing that H were true.

There are two great figures in the philosophy of probability associated with this view. One is Frank P. Ramsey, the other one is the Italian statistician Bruno de Finetti (1972, 2008). Since Ramsey's work is prior to de Finetti's and also influenced the latter's views, we shall focus on Ramsey. Here is his famous analysis of conditional degrees of belief:

If two people are arguing 'if H will E?' and both are in doubt as to H, they are adding H hypothetically to their stock of knowledge and arguing on that basis about E. (Ramsey, 1926)

The above quote is to a certain extent ambiguous: it is about conditional (degree of) belief or about the assessment of the truth or probability of conditionals? Many philosophers, most famously Stalnaker (1968, 1975), were inspired by the second reading and developed a theory of (the probability of) conditionals based on the idea that assessing the conditional $H \rightarrow E$ involves a counterfactual addition of H one's background knowledge.

I would like to stay neutral on all issues concerning conditionals and interpret Ramsey's quote as an analysis of conditional degrees of belief. Indeed, in the sentence that follows the above quote, Ramsey describes the entire procedure as

We can say that they are fixing their degrees of belief in E given H. (ibid.)

This makes clear that regardless of the possible link to the probability of conditionals, Ramsey intended that hypothetically assuming H would determine one's conditional degrees of belief in E, given H. That is, $p(E|H)$ is the rational degree of belief in E if we supposed that H were true. This analysis directly yields The Equality: if we *suppose* that H is true and the coin is fair, then all events in this world ω_H are described by the probability density $\rho_H(E)$. The fictitious world ω_H is genuinely chancy—and we happen to know what these chances are. If a chance-credence calibration norm is ever to work (Strevens, 1999), this must be the place: our degrees of belief (conditional on

supposing H) should follow the objective probabilities, that is, the probability densities given by H.

It is important to note the difference to the Principal Principle, already pointed out in the introduction. That principle applies to real-world, ontic chances, e.g., “the chance of this atom decaying in the next hour is $1/3$ ” or “the chance of a zero in the spin of this roulette wheel is $1/37$ ”. The Principal Principle simply claims that degrees of belief should mirror such chances. Compare this to the picture that we sketch for conditional degree of belief. We are not interested in real-world chances; rather we observe that *in the world ω_H described by H*, there is an objective and unique chance of E occurring, and it is described by the probability density $\rho_H(E)$. This is just what it means to suppose that H is the case. In other words, we apply the Principal Principle not in the real world, but in the counterfactual world where H holds, and we adapt our (conditional) degree of belief in E to $\rho_H(E)$. By supposing a world where the occurrence of E is genuinely chancy, the Ramseyian account of conditional degree of belief explains why our conditional degree of belief in E given H is uniquely determined and obeys The Equality.

In other words, Bayesian inference requires two things for coordinating conditional degrees of belief with probability densities: First, the Ramseyian, counterfactual interpretation of conditional degree of belief which entails the supposition that H is true—even if actually know that it is false. Second, conditional on this supposition, it is arguably a requirement of rationality to coordinate degrees of beliefs with known objective chances (Lewis, 1980). If this coordination is denied (as Strevens, 1999, seems to do when he points to a “justification gap”), then degrees of belief are never rationally constrained by chances and all Bayesian inference has to be radically subjective. But apart from Strevens himself, nobody seems willing to draw this conclusion.

To repeat, we are talking about chance-credence coordination in hypothetical worlds where the space of possible events (=the sampling space) is very restricted, not about chance-credence coordination in the actual world $\omega_@$. This consequence is very desirable because most statistical models are strong idealizations of the real world that neither capture physical propensities, nor limiting frequencies, nor chances according to a best-system account. Think of the omnipresent assumption of normality of errors, focusing on specific causally relevant factors and leaving out others, and so on. Probability densities in statistics can only inform our *hypothetical* degrees of belief, not our actual degrees of belief. However, this is exactly what we need to calculate posterior probabilities and (Bayesian) measures of evidence, such as the Bayes factor.

Incidentally, this interpretation of conditional degree of belief fits well with the

thoughts of the great (non-Bayesian) statistician Ronald A. Fisher on the nature of conditional probability in statistical inference:

In general tests of significance are based on *hypothetical* probabilities calculated from their null hypotheses. They *do not lead to any probability statements about the real world*. (Fisher, 1956, 44, original emphasis)

That is, Fisher is emphatic that the probabilities of evidence given some hypothesis have hypothetical character and are not objective chances with physical reality. According to Fisher, statistical reasoning and hypothesis testing is essentially counterfactual—it is about the probability of a certain dataset under the tested “null” hypothesis. The null hypothesis usually denotes the absence of any effect, the additivity of two factors, the causal independence of two variables in a model, etc. In most cases, it is strictly speaking false: there will be *some* minimal effect in the treatment, some slight deviation from additivity, some negligible causal interaction between the variables. Our statistical procedures are thus based on the probabilities of events under a hypothesis which we know to be false—although it may be a good idealization of reality (Gallistel, 2009). Hence, the proposed counterfactual interpretation of conditional degree of belief naturally fits into the practice of statistical inference.

An additional advantage of this strategy is that we do not have to choose between different formulations of the Principal Principle, e.g., regarding the problem of admissible information (Lewis, 1980). That problem consisted in a conflict between externally given information related to an event and the proper physical chance of that event in a well-defined experiment. However, the worlds ω_H are so simple and well-behaved—H assigns a probability to all events in the sampling space—that these conflicts cannot occur. Our actual background knowledge cannot affect the conditional degrees of belief because we take a counterfactual stance, suppose that H were the case and screen off any conflicting information. We can thus explain why chance-credence coordination is so important to probabilistic reasoning, without committing ourselves to the tedious task of defining and refining the Principal Principle in the complex and messy actual world $\omega_{@}$.

4 Implications for Bayesian Inference

We now turn to the implications of the Ramseyian approach to conditional degree of belief for Bayesian inference and statistical reasoning in general.

First, it may seem that the statistical hypothesis H and the background assumptions are not clearly demarcated. Consider the case of tossing a coin. When we evaluate $p(E|H)$ with $H = \text{“the coin is fair”}$, we typically assume that the individual tosses of the coin are independent and identically distributed. However, this assumption is typically not part of H itself. If we contrast H to some alternative H' , we notice that the differences between them are exclusively expressed in terms of parameter values, such as $H: \theta = 1/2$ versus $H': \theta = 2/3$, $H'': \theta > 1/2$, etc. So it seems that assumptions on the experiment, such as independence and identical distribution of the coin tosses, do not enter the particular hypothesis we are testing. Rather, they are part of general statistical model in which we compare H to H' . In other words, there are two layers in the statistical modeling process—the layer of the general experimental model $\mathcal{M} = (B(N, \theta), \theta \in [0, 1])$ which contains the assumptions of independence and identical distribution as well as the modeling of the individual toss as a Bernoulli (success/failure) experiment, and the layer of the particular hypotheses about the value of θ , that is, the probability of the coin to come up heads.

This implies that the conditional degree of belief $p(E|H)$ is not only conditional on H , but also conditional on \mathcal{M} . Indeed, the typical Bayesian inference about the probability of heads in the coin-tossing example takes \mathcal{M} as given from the very start. Hence, also the assignment of prior probabilities $p(H)$ takes \mathcal{M} as given: a Bayesian distributes her degrees of belief only over elements of \mathcal{M} . Bayesian inference regarding particular parameter values is relative to a model into which all hypotheses are embedded. In particular, also the prior and posterior degrees of belief, $p(H)$ and $p(H|E)$, should be understood as relative to a model \mathcal{M} .

This move resolves a simple, but pertinent problem of Bayesian inference. On the subjective interpretation of probability, the probability of a proposition H , $p(H)$, is standardly interpreted as the degree of belief that H is true. However, in science, we are often in a situation where we know that all of our models are strong idealizations of reality and where we would not have a strictly positive degree of belief in the literal truth of a certain statistical hypothesis. Similarly, the outcome space is highly idealized: a coin may end up balancing on the fringe, a toss may fail to be recorded, the coin may be destroyed, etc. All these possibilities have a certain probability, but we neglect them when setting up a statistical model and interpreting an experiment.

In other words, Bayesian inference seems to be based on false and unrealistic premises: the interpretation of degrees of belief that H is *true* fails to make sense for $p(H)$. So how can Bayesian inference ever inspire confidence in a hypothesis? Do we have to delve into the muddy waters of approximate truth, verisimilitude, and so on?

No. The considerations in this paper suggest a much simpler alternative: to interpret prior probabilities as *conditional* (and counterfactual) degrees of belief, that is, degrees of belief in H that we would have if we supposed that the general model of the experiment \mathcal{M} were true. This move solves this problem by making the entire Bayesian inference relative to \mathcal{M} . And the adequacy of \mathcal{M} is a matter of external judgment and not of reasoning within the model. That is, a Bayesian inference is trustworthy to the extent that the underlying statistical model is well-chosen and the prior probabilities are well motivated. Of course, this is no peculiar feature of Bayesian inference: it is characteristic of all scientific modeling. Garbage in, garbage out. We have thus answered the question of why Bayesian inference makes sense if all statistical models are known to be wrong, but some are illuminating and useful (Box, 1976). More generally, we have assimilated statistical reasoning to other ways of model-based reasoning in science (e.g., Weisberg, 2007; Frigg and Hartmann, 2012).

This account of conditional degree of belief squares well with a Humean account of objective chance, where chances are stable, but imperfect regularities in a description of a target system that optimally balances simplicity, strength and fit (Lewis, 1994; Hoefer, 2007). Note, however, that our account is compatible with any analysis of objective chances in the real world. It is a distinct strength of the analysis presented here that it remains neutral on the nature of objective chance: scientists and statisticians do not have to engage in metaphysical analyses of objective chance when they build a statistical model and use it to inform their credences.

We also notice that not all conditional degrees of belief are of the same kind. There is a relevant difference between $p(E|H, \mathcal{M})$ on the one hand and $p(H|\mathcal{M}, E)$ on the other hand. When we calculate the first value, we suppose that \mathcal{M} and H are the case and argue that the probability of E should be equal to $\rho_{\mathcal{M}, H}(E)$. The second value, however, is commonly interpreted as the probability of H after *learning* E, given \mathcal{M} . Indeed, supposing \mathcal{M} and E does not yield a uniquely rational value for $p(H|\mathcal{M}, E)$. There is no objective chance of H in the hypothetical world ω_E . The only way to argue that $p(H|\mathcal{M}, E)$ is objectively determined is to make use of Bayes' Theorem: the already determined values of $p(E|H, \mathcal{M})$, $p(E|\mathcal{M}, H)$ and $p(H, \mathcal{M})$ force us to assign a particular value to $p(H|\mathcal{M}, E)$ if we want to stay coherent.

This suggests that the *learning* of evidence E should be modeled differently from supposing E, again in agreement with recent psychological evidence that points to differences between both modes of reasoning (Zhao et al., 2012). The obvious option is to state that the posterior probability of H after learning E should be modeled by the probability $p^E(H|\mathcal{M})$ which represents the agent's rational degree of belief after learn-

ing E and should be numerically identical to (but not be counterfactually interpreted as) the conditional probability $p(H|E, \mathcal{M}) = p(H|\mathcal{M}) \cdot p(E|H, \mathcal{M}) / p(E|\mathcal{M})$. In other words, $p^E(H|\mathcal{M})$ should always correspond to the value of $p(E|H, \mathcal{M})$, but not be counterfactually interpreted as the rational degree of belief in H if we supposed that \mathcal{M} and E were the case. Writing $p^E(H|\mathcal{M})$, which represents a diachronic shift to another probability distribution, as $p(H|E, \mathcal{M})$, a conditional probability in the original distribution, is an abuse of notation, strictly speaking.

This proposal resembles the treatment of conditional probability which we have endorsed following Hájek (2003): Ratio Analysis provides a numerical *constraint* on conditional probability without providing a *definition* or exhaustive analysis. Analogously, Bayes' Theorem does not provide a definition of $p^E(H|\mathcal{M})$, but a constraint on these probabilities. The rational posterior degree of belief in H given \mathcal{M} and after learning E should just be the number yielded by applying Bayes' Theorem to $p(H|E, \mathcal{M})$. Both learning and supposing are indispensable for Bayesian inference, and both modes of reasoning are exemplified in the probabilities $p^E(H|\mathcal{M})$ and $p(E|H, \mathcal{M})$. Note that we do not need to argue why the posterior probability $p^E(H|\mathcal{M})$ should equal $p(E|H, \mathcal{M})$. This is just how Bayesians model the learning of evidence, and as such, this question is orthogonal to the scope of this paper.

Making the learning-supposing difference explicit also allows for a better treatment of the Problem of Old Evidence. For arguing that "old" (=previously known) evidence confirms a hypothesis H , most solutions make use of an argument of the form $p(E|H) \gg p(E|\neg H)$ (e.g., Garber, 1983; Howson, 1984; Earman, 1992; Sprenger, 2015; Fitelson and Hartmann, 2016). When conditional degree of belief is interpreted in the standard way, according to Ratio Analysis, these values are equal to unity. When we evaluate conditional degree of belief by supposing the hypothesis, however, one can meaningfully state that old evidence supports the hypothesis in question. Hence, this approach backs up the proposed technical solutions by a philosophical story of why it is reasonable to have non-trivial conditional degrees of belief. Similarly, it supports those Bayesians who believe that Bayes factors and posterior probabilities can be objective—or at least intersubjectively compelling—measures of evidence (Sprenger, 2016).

The consequence of all this is that—at least for the purpose of statistical reasoning and scientific inference—conditional and not unconditional degree of belief should be taken as a primitive notion. This resonates well with Hájek's (2003) analysis which reaches the same conclusion. It requires some changes on the axiom level, however. Kolmogorov's three standard axioms ($p(\perp) = 0$; $p(A) + p(\neg A) = 1$;

$p(\bigvee A_i) = \sum p(A_i)$ for mutually exclusive A_i) will not do any more. One way is to replace them by an axiom system that takes conditional probability as primitive, such as the Popper-Renyi axioms (Renyi, 1970; Popper, 2002). Another way is to define conditional probability not in terms of unconditional probability, as in Ratio Analysis, but more generally in terms of an expectation conditional on a random variable (Gyenis et al., 2015). Personally, I prefer the first option because the gain in simplicity by moving to conditional probability as a primitive notion comes at no loss of mathematical strength.

5 Conclusion

This paper was devoted to a defense of why conditional degrees of belief should equal the corresponding probability densities, as The Equality postulated. We can now state our results. Probably, many philosophers of probability share these views, but I am aware of no place where they are made explicit and defended.

1. Ratio Analysis is an inadequate explication of conditional degree of belief.
2. Conditional degrees of belief $p(E|H)$ should be interpreted in the counterfactual way outlined by Ramsey.
3. If this approach is chosen in statistical inference, then supposing H will uniquely determine the objective probability of E and therefore (by chance-credence coordination) also the conditional degree of belief in E. In other words, we have justified The Equality.
4. This approach avoids problems with more demanding principles of chance-credence coordination such as the Principal Principle.
5. It explains the seemingly analytical nature of many probability statements in statistics, and it agrees with the view of statisticians on probability in inference: namely as hypothetical entities.
6. It gives a philosophical foundation to classical and recent solutions of the Problem of Old Evidence and makes measures of evidence in Bayesian statistics more objective. In other words, it rescues the normative force of Bayesian inference.
7. The Ramseyian interpretation also transfers to prior probabilities in a Bayesian model which are conditional on assuming a general statistical model. This an-

swers the challenge why we should ever have positive degrees of belief in a hypothesis when we know that the model is wrong.

8. However, not all degrees of belief in statistical inference should be evaluated counterfactually: learning evidence is radically different from supposing it, even if the probabilities agree numerically.

If only some of these conclusions withstood the test of time, that would already be a fair result.

References

- Box, G. E. P. (1976). Science and Statistics. *Journal of the American Statistical Association*, 71:791–799.
- de Finetti, B. (1972). *Probability, Induction and Statistics: the Art of Guessing*. John Wiley & Sons, New York.
- de Finetti, B. (2008). *Philosophical Lectures on Probability*. Springer, Dordrecht, Holland.
- Earman, J. (1992). *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. MIT Press, Cambridge, Mass.
- Fisher, R. (1956). *Statistical methods and scientific inference*. Hafner, New York.
- Fitelson, B. and Hájek, A. (2016). Declarations of independence. *Synthese*.
- Fitelson, B. and Hartmann, S. (2016). A New Garber-Style Solution to the Problem of Old Evidence. *Philosophy of Science*.
- Frigg, R. and Hartmann, S. (2012). Models in Science. *Stanford Encyclopedia of Philosophy*.
- Gaifman, H. and Snir, M. (1982). Probabilities Over Rich Languages, Testing and Randomness. *The Journal of Symbolic Logic*, 47(3):495–548.
- Gallistel, C. R. (2009). The importance of proving the null. *Psychological review*, 116(2):439–453.
- Garber, D. (1983). Old evidence and logical omniscience in Bayesian confirmation theory Geciteerd door mij. *Testing scientific theories*.

- Gyenis, Z., Rédei, M., and Hofer-Szabó, G. (2015). Conditioning using conditional expectations: The Borel-Kolmogorov Paradox.
- Hájek, A. (2003). What Conditional Probability Could Not Be. *Synthese*, 137:273–323.
- Hoefer, C. (2007). The Third Way on Objective Probability: A Sceptic’s Guide to Objective Chance. *Mind*, 116(463):549–596.
- Howson, C. (1984). Bayesianism and support by novel facts. *British Journal for the Philosophy of Science*, pages 245–251.
- Howson, C. (1991). The ‘Old Evidence’ Problem. *British Journal for the Philosophy of Science*, 42:547–555.
- Howson, C. and Urbach, P. (2006). *Scientific Reasoning: The Bayesian Approach*. Open Court, La Salle, IL, 3rd edition.
- Jackson, F. (1991). *Conditionals*. Oxford University Press, Oxford.
- Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90:773–795.
- Lewis, D. (1980). A Subjectivist’s Guide to Objective Chance. In Jeffrey, R. C., editor, *Studies in Inductive Logic and Probability*, pages 263–293. University of California Press, Berkeley.
- Lewis, D. (1994). Humean Supervenience Debugged. *Mind*, 103:473–490.
- Popper, K. R. (2002). *The Logic of Scientific Discovery*. Routledge, London, reprint edition.
- Ramsey, F. P. (1926). Truth and Probability. In Mellor, D. H., editor, *Philosophical Papers*, pages 52–94. Cambridge University Press, Cambridge.
- Renyi, A. (1970). *Foundations of Probability*. Holden-Day Inc., San Francisco.
- Rosenthal, J. (2004). *Wahrscheinlichkeiten als Tendenzen*. mentis, Paderborn.
- Savage, L. J. (1972). *The Foundations of Statistics*. Dover, New York, 2nd edition.
- Skyrms, B. (2000). *Choice & Chance*. Wadsworth, Belmont, CA, 4th edition.

- Sprenger, J. (2010). Statistical inference without frequentist justifications. In *EPSA Epistemology and Methodology of Science: Launch of the European Philosophy of Science Association*, pages 289–297. Springer Netherlands.
- Sprenger, J. (2015). A Novel Solution of the Problem of Old Evidence. *Philosophy of Science*.
- Sprenger, J. (2016). The Objectivity of Subjective Bayesian Inference.
- Stalnaker, R. (1968). A Theory of Conditionals. In *Studies in Logical Theory: American Philosophical Quarterly Monograph Series, No. 2*. Blackwell, Oxford.
- Stalnaker, R. (1975). Indicative Conditionals. *Philosophia*, 5(3):269–286.
- Strevens, M. (1999). Objective Probability as a Guide to the World. *Philosophical Studies*, 95:243–275.
- Suarez, M. (2011a). *Probabilities, Causes and Propensities in Physics*. Springer, New York.
- Suarez, M. (2011b). Propensities and Pragmatism. *Journal of Philosophy*, 100:61–92.
- Weisberg, M. (2007). Who is a Modeler? *The British Journal for the Philosophy of Science*, 58:207–233.
- Williamson, J. (2007). Motivating Objective Bayesianism: From Empirical Constraints to Objective Probabilities. In Harper, W. and Wheeler, G., editors, *Probability and Inference: Essays in Honour of Henry E. Kyburg Jr.*, pages 151–179. College Publications, London.
- Williamson, J. (2010). *In Defence of Objective Bayesianism*. Oxford University Press, Oxford.
- Zhao, J., Crupi, V., Tentori, K., Fitelson, B., and Osherson, D. (2012). Updating: Learning versus Supposing. *Cognition*, 124:373—378.
- Zhao, J., Shah, A., and Osherson, D. (2009). On the provenance of judgments of conditional probability. *Cognition*, 113(1):26–36.