Conditional Degree of Belief

Jan Sprenger*

February 24, 2016

1 Introduction

Bayesian inference is based on representing degrees of belief by probabilities, and on changing these degrees of belief by Conditionalization. Its normative force in science relies to a large extent on constraints on the conditional degree of belief in evidence E given hypothesis H, that is, $p(E|H)$.

To motivate this claim, think of what Bayesian inference in science typically looks like. There are competing statistical hypotheses H_θ, $\theta \in [0,1]$, that model a physical process, such as the chance of a coin to come up heads. Each of the H_θ specifies the probability of a certain event E (e.g., two heads in three tosses) by means of a mathematical function, the *probability density* $\rho_{H_\theta}(E)$. 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.). The corresponding probability density is $\rho_{H_\theta}(E) = \binom{N}{k}\theta^k(1-\theta)^{N-k}$. When, for any particular H (e.g., $\theta = 1/2$), the conditional degree of belief $p(E|H)$ tracks the probability density, that is, $p(E|H) = \rho_H(E)$, a Bayesian uses this value for calculating the posterior degree of belief in H.

The equality $p(E|H) = \rho_H(E)$ is important for a variety of reasons. First, it allows to calculate the Bayesian's primary measure of evidential support, the *Bayes factor* (Kass and Raftery, 1995). Bayes factors quantify 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 are unconstrained, it follows that one may not be able to agree whether E favors H_0 over H_1, or vice

*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

versa. For the application of Bayesian inference in the sciences, that would be a heavy 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 and more generally, according to the Likelihood Principle (Birnbaum, 1962; Berger and Wolpert, 1984), which is accepted by Bayesians as well as by some non-Bayesians (Hacking, 1965; Sober, 2008), all experimental evidence that E yields about an unknown parameter θ is contained in the values of the *likelihood function* $L(\theta) := p(E|H_\theta)$ for different values of θ. Without agreement on these values, the evidential relevance of the observations remains unclear and contested, with devastating consequences for the epistemic authority of scientific research based on Bayesian methods.

Third, 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., Blackwell and Dubins, 1962; 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 ¬H is true. However, if the degrees of belief in E given H (or ¬H) vary, agents will also reach different posterior probabilities in the limit. This undermines the trustworthiness of Bayesian inference and the decision-theoretic paradigm that posterior probabilities should guide our decisions (Savage, 1972).

Hence, the main question of our paper is: What justifies the equality between conditional degrees of belief and probability densities that is required for meaningful Bayesian inference?

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

Various scholars derive The Equality from a general epistemic norm. They consider it 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, according to the *Principle of Direct Inference (PDI)* (e.g., Reichenbach, 1949; Kyburg, 1974; Levi, 1977), if I know that a coin is fair, I should assign degree of belief 1/2 that heads will come up. David Lewis (1980) formalized a related intuition in his *Principal Principle (PP)*: the initial credence function of a rational agent, conditional on the proposition that the physical chance of E takes value x, should also be equal to x.

Do chances in statistical inference conform to those descriptions? Partially, they do. A statistical hypothesis H describes the probability of possible events E by the probability density $\rho_H(E)$. So the chance of E given H is 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 $\rho_H(E) = 3/8$.

However, this probability is objective, but 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, nor about the long-run frequency of E. Instead, it describes the objective chance of E *if the statistical hypothesis 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. Indeed, many scientific models are of this idealized type and do not have any real-world implementation. The Principle of Direct Inference and the Principal Principle, by contrast, apply to physical chances: there has to be an event whose objective chance is 3/8. They are silent on chances that are conditional on an idealized statistical model. Therefore, they fail to justify The Equality. It is the task of this paper to provide such a justification.

Hence, our paper joins the efforts by philosophers of science to clarify the nature of objective probability in scientific reasoning. Hoefer (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. (Hoefer, 2007, 550)

While the role of physical chance in explanation and prediction is well-explored (Hoefer, 2007; Suarez, 2011a,b), the role of mathematical probability densities and their relation to conditional degree of belief is much less explored.

The rest of the paper is structured as follows. Section 2 examines the standard approach to conditional degree of belief: the ratio analysis of conditional probability. I argue that it fails to account for The Equality, and also as an explanation of our probabilistic reasoning practices. 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 is devoted to exploring the implications of our proposal for Bayesian inference: the counterfactual interpretation of conditional degree of belief should not be restricted to likelihoods, but also hold for prior probabilities. This move explains why Bayesian reasoners may assign positive degrees of belief to a hypothesis which they know to be wrong. Section 5 discusses the interpretation of posterior degrees of belief and the different roles of learning and supposing in Bayesian inference. Section 6 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).

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

Now, there seems to be a fast and easy argument why the ratio analysis of conditional probability fails to cover many important cases of Bayesian inference. Suppose that we reason about the bias of a coin, with H_θ denoting the hypothesis that the bias of the coin is equal to $\theta \in [0, 1]$. If our prior probability density over the H_θ 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 conditional degree of belief 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 since intuitively, we are able to assign a probability to the event of E given H (=the conditional event E | H)—both from a frequentist and a subjective Bayesian perspective (de Finetti, 1972, 81).

A similar challenge is presented by Borel's paradox and related puzzles: "What is the probability that a point on Earth is in the Western hemisphere (H), given that it lies on the equator (E)?". Intuitively, there seems to be a rational answer (one half), but according to Ratio Analysis, we cannot even express such a degree of belief because for a uniform probability distribution, $p(E) = 0$. Notably, even the well-known mathematics textbook by Billingsley (1995, 427) writes that "the conditional probability of [H] with respect to [E, ...] is defined [by Ratio Analysis] unless $p(E)$ vanishes, in which case it is not defined at all" (notation changed). In other words, the epistemic function of conditional degree of belief is not adequately mirrored in the mathematics of conditional probability. The seminal article by Alan Hájek (2003)—"What Conditional Probability Could Not Be"—discusses these problems of Ratio Analysis in detail.

However, Hájek also mentions a refinement of Ratio Analysis that is based on representing events and hypotheses as measurable sets. Let $\mathcal{H}$ denote a measurable set of hypotheses (e.g., $\theta \in \Theta$). Then a conditional probability $p(\cdot|\mathcal{H})$ is a measurable function from the σ-algebra of possible events to the real numbers such that

$$p(E\mathcal{H}) = \int_{\mathcal{H}} p(E|\mathcal{H})dp \qquad \text{(Refined Ratio Analysis)}$$

This definition of conditional probability, due to Kolmogorov (1933), is implicit: conditional probabilities are functions that yield the product event $p(E\mathcal{H})$ if one integrates them over $\mathcal{H}$, relative to the probability measure $p(\cdot)$. For atomic hypothesis sets $\mathcal{H} = \{H\}$, Refined Ratio Analysis reduces to Ratio Analysis, as one can readily verify. Howson and Urbach (e.g., 2006, 37) propose a similar approach for parametric statistical inference: if $p(H_\theta) = 0$ for a specific parameter value θ, then we define $p(E|H_\theta)$ as the limit of $p(E|(\theta - \varepsilon, \theta + \varepsilon))$ for $\varepsilon \to 0$. As long as the prior probability density over θ is continuous and strictly positive, this quantity will be well-defined and can act as a surrogate for the direct calculation of $p(E|H_\theta)$.

Easwaran (2011) and Myrvold (2015) show how this approach can be used to answer Borel's paradox and to rescue statistical inference with likelihood functions. For the sake of the argument, I shall not try to take a stand in the debate and just concede that Refined Ratio Analysis may be sufficient to talk meaningfully about events with probability zero (though see Fitelson and Hájek (2016) for a dissenting view). The focus of my argument is different: even if the technical problems are solved, Ratio Analysis fails to give an adequate explanation of our cognitive practices. It also fails to account for The Equality and to explain why conditional probabilities are often constrained in a seemingly objective way. For reasons of simplicity, I shall focus on those cases where probability zero is not a problem and Refined Ratio Analysis reduces to Ratio Analysis.

I shall begin with the first objection. Ratio Analysis defines conditional degree of belief in terms of unconditional degree of belief. The advantage of this move is that we need no separate conceptual analysis of conditional degree of belief. Neither do we need a bridge between probability densities and conditional degrees of belief: the latter are reduced to unconditional degrees of belief. However, this approach fails to do justice to the cognitive role of conditional degrees of belief. We do not 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. Indeed, recent psychological evidence suggests that Ratio Analysis is a poor description of how people reason with conditional probabilities, pointing out the necessity of finding an alternative account (Zhao et al., 2009).

Second, Ratio Analysis fails to grasp the normative role of conditional degree of belief in statistical inference. In the introduction, we have seen that it is part of the *meaning* of H to constrain $p(E|H)$ in a unique way. Recall the example. 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*. On Ratio Analysis, this feature of conditional degree of belief drops out of the picture. $p(E|H)$ is constrained only via constraints on $p(E \wedge H)$ and $p(H)$. But even if we suspend judgment on $p(E \wedge H)$ and $p(H)$, we may still have a definite degree of belief $p(E|H)$. Ratio Analysis therefore misses an important aspect of conditional degree of belief. More precisely, as long as the overall probability distribution is coherent, we do not have to coordinate probability densities and conditional degrees of belief according to Ratio Analysis. While this result may be acceptable for radically subjective Bayesians, it is devastating for those philosophers (like myself) who believe that Bayesian inference should be constrained by the implications of statistical hypotheses for the probabilities of real-world events.

Finally, it is notable that Refined Ratio Analysis *defines* a new conditional density rather than calculating the value of the density via Ratio Analysis. That is, to the extent that it circumvents the problems created by the naïve definition of Ratio Analysis, it acknowledges the need of viewing conditional probability as an independent, and possibly primitive, concept. The Howson-Urbach definition where conditional probability is calculated via a mathematical limit is a particularly salient example. Unsurprisingly, Fitelson and Hájek (2016) advocate replacing Kolmogorov's axioms by the Popper-Renyi axioms (Renyi, 1970; Popper, 2002) where conditional probability is take as primitive. This is a view that matches my own inclinations well, and the following section will show how it naturally emerges from a proper understanding of conditional degree of belief.

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). I will focus on Ramsey since de Finetti also requires that H must be a (verifiable) event for $p(E|H)$ to be meaningful (de Finetti, 1972, 193). This is unnecessarily restrictive.

Here is Ramsey's 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 ambiguous: it is about conditional (degree of) belief, the truth conditions of conditionals, or about their probability? Many philosophers, most famously Stalnaker (1968, 1975), were inspired by the latter readings and developed a theory of (the probability of) conditionals based on the idea that assessing the conditional H→E involves adding H to one's background knowledge. See also Douven (2016).

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., my emphasis)

This makes clear that regardless of the possible link to the epistemology 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, then all events E are described by the probability density $\rho_H(E)$. The fictitious world ω_H, created by supposing H, is genuinely chancy, regardless of whether the actual world is. Indeed, it can be seen as part of the *meaning* of $\rho_H(E)$ to quantify the chances of E (and other events) if H were the case. If a chance-credence calibration norm is ever to work, this must the place: our degrees of belief, conditional on supposing H, should follow the objective chances that H imposes on other events.

It is important to note the difference to the Principle of Direct Inference (PDI) and the Principal Principle (PP), already pointed out in the introduction. Both principles apply 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 principles simply claim 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 PDI/PP 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. Note that this is really an initial credence function, as PP demands: information about the actual world $\omega_@$ that may conflict with H is irrelevant in ω_H.

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 supposes that H is true— even if we 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, few are 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.

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 physically realized objective chances. Probabilities are useful instruments of inference, not components of the actual world. 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 with its emphasis on testing idealized point hypotheses, e.g., in null hypothesis significance testing.

Finally, I would like to point out that problems with inadmissible information, that affect the Principal Principle, do not occur in this proposal. 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 for probabilistic reasoning, without committing ourselves to the tedious task of transferring the Principal Principle from initial to actual credences in the world $\omega_{@}$.

4 Implications I: The Model-Relativity of 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 = "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 statistical model $\mathcal{M} = (\mathcal{S}; \mathcal{P})$ where $\mathcal{S}$ denotes the sampling space, that is, the set of observations, and $\mathcal{P}$ denotes the set of probability distributions that describe the observations. In the case of i.i.d. coin tosses with fixed sample size N, this would be the set of Binomial distributions $B(N, \theta)$, with $\theta \in [0, 1]$ denoting the probability of the coin to come up heads on any particular toss.

This implies that the conditional degree of belief $p(E|H)$ is not only conditional on H, but also conditional on $\mathcal{M}$. Indeed, a Bayesian inference about the probability of heads in the coin-tossing example takes $\mathcal{M}$ as given from the very start. This is especially clear in the assignment of prior probabilities $p(H)$: Bayesian inference regarding particular parameter values is relative to a model $\mathcal{M}$ into which all hypotheses are embedded, and degrees of belief are distributed only over elements of $\mathcal{M}$. In particular, also the prior and posterior degrees of belief, $p(H)$ and $p(H|E)$, should be understood as relative to a model $\mathcal{M}$. In the above example, a Bayesian might distribute her prior beliefs according to a probability density over $\theta \in [0, 1]$, such as the popular beta distribution Beta(α, β) with density $p(\theta \in [a, b]) = \int_a^b (1/B(\alpha, \beta))\, x^{\alpha-1}\, (1-x)^{\beta-1}\, dx$.

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. It would be silly to have a strictly positive degree of belief in the 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 damaged, etc. All these possibilites 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. Instead of $p(H)$, we talk about $p(H|\mathcal{M})$. This move solves this problem by making the entire Bayesian inference relative to $\mathcal{M}$. Similarly, we replace the marginal likelihood $p(E)$ by $p(E|\mathcal{M})$ and interpret it counterfactually, in agreement with the Law of Total Probability. $p(E|\mathcal{M})$ is the weighted average of the conditional probabilities of E, and thus our subjective expectation that E occurs if $\mathcal{M}$ were the case.

$$p(E) = \sum_{H \in \mathcal{M}} p(E|H, \mathcal{M}) \cdot p(H|\mathcal{M})$$

On this view, the adequacy of $\mathcal{M}$ becomes 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).

Note that this 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. This is the same intuition that motivated the best-system analysis of objective chance (Lewis, 1994; Hoefer, 2007).

The model-relativity of a lot of probabilistic inference, as well as the results of the last two sections, suggest that conditional and not unconditional degree of belief might be a more adequate primitive notion if one wants to keep a sparse conceptual framework for probabilistic reasoning. 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 ($p(\bot) = 0$; $p(A) + p(\neg A) = 1$; $p(\vee A_i) = \sum p(A_i)$ for mutually exclusive A_i) will not do the job 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; Fitelson and Hájek, 2016). Unconditional probability can then be obtained as a limiting case of conditional probability. Another way is to define conditional probability in terms of an expectation conditional on a random variable (Gyenis et al., 2016). It is up to future work to determine which road is the most promising one.

5 Implications II: Learning vs. Supposing

Following up on the thoughts in the previous section, we 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 $p(H|E, \mathcal{M})$ 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 $\rho_{\mathcal{M},H}(E)$. However, 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. On the other hand, $p(H|E, \mathcal{M})$ is objectively constrained by Bayes' Theorem: the already determined values of $p(E|H, \mathcal{M})$ $p(E|\mathcal{M})$ and $p(H, \mathcal{M})$ force us to assign a particular value to $p(H|\mathcal{M}, E)$ if we want to stay coherent.

In other words, our Ramseyian analysis of conditional degree of belief seems to be very one-sided: it is restricted to likelihoods (probabilities of events given a specific statistical distribution). For these probabilities of the type $p(E|H, \mathcal{M})$, supposing H and $\mathcal{M}$ delivers the intuitive result that the conditional degrees of belief are objectively constrained, but supposing E and $\mathcal{M}$ does not deliver such a result for $p(H|E, \mathcal{M})$. Given that Bayesians interpret this value as the rational posterior degree of belief in H and take it as a basis for decision-making and action, this seems to wreak havoc on our attempts to integrate the Ramseyian approach to counterfactual degree of belief into an appealing picture of Bayesian reasoning.

Perhaps contraintuitively, I suggest to bite the bullet and to accept that the conditional degree of belief $p(H|E, \mathcal{M})$ is unconstrained in the first place. Supposing E and

$\mathcal{M}$ typically allows for different degrees of belief in H. There is no reason why, in the absence of further information, agents should assign the same degree of belief to H, given E. In fact, they often don't. Now assume that we have a definite degrees of belief in $p(H|\mathcal{M})$ (as a matter of psychological fact) and $p(E| \pm H, \mathcal{M})$ (by supposing H and ¬H, as explained in Section 3). Then our degrees of belief in $p(E|\mathcal{M})$ and $p(H|E, \mathcal{M})$ are fixed as well because $p(\cdot|\mathcal{M})$ is a probability function and therefore satisfies

$$p(H|E, \mathcal{M}) = p(H|\mathcal{M}) \cdot \frac{p(E|H, \mathcal{M})}{p(E|\mathcal{M})} \qquad \text{(Bayes' Theorem)}$$

If our conditional degrees of belief do not satisfy this equality, we will run into a Dutch Book and violate the Bayesian norms of rationality. Three things should be noted, however.

First, the left hand side on the equation describes the conditional degree of belief in H, given E and $\mathcal{M}$, not on the degree of belief in H, given $\mathcal{M}$ and after learning E. (I will say more on this distinction below.) It is a synchronic, not a diachronic constraint, and Bayesian Conditionalization has not been invoked in stating the above equality.

Second, Bayes' Theorem is a mathematical fact which is in itself closely related to Ratio Analysis—in fact, it can be derived easily from applying Ratio Analysis to both $p(E|H, \mathcal{M})$ and $p(H|E, \mathcal{M})$. This underlines that the above equation is not about the *definition* of $p(H|E, \mathcal{M})$, but about constraining its value.

Third and last, we see that $p(H|E, \mathcal{M})$ is constrained to the extent that the probabilities on the right hand side are. In other words, the normative pull for constraining $p(H|E, \mathcal{M})$ does not come from the counterfactual interpretation of conditional degrees of belief (=supposing E), but from the requirement of probabilistic coherence. This is a reading of Bayesian inference that is very close to Ramsey's and de Finetti's subjectivism: we are free to assign degrees of belief to events as long as we respect probabilistic coherence and the constraints which rational arguments impose on us (e.g., when we suppose that H is the case and reason about the probability of E). This clarifies that subjectivism is not to be confused with an "anything goes" mentality: the rationally required constraints can be substantial.

It should also be added that supposing a certain event E may still determine the conditional probability of other events. Suppose that two variables $X = (x_t)$ and $Y = (y_t)$ depend on each other via the formula $y_{t+1} = \alpha x_{t+1} + \beta y_t + \varepsilon_{t+1}$, a typical time series model $\mathcal{M}$ with Gaussian error terms $\varepsilon \sim N(0, 1)$. Let H denote specific values for α and β, e.g., $\alpha = \beta = 1$. In that case, the conditional degree of belief in $E' : y_{t+1} = 2$ given H, $p(E'|H, \mathcal{M})$, is not uniquely determined. However, if we also suppose that $E : y_t = 1$ occurred, the model yields a meaningful and uniquely counterfactual degree of belief $p(E'|H, E, \mathcal{M})$.

Finally, inference about posterior degrees of belief typically concerns the case when we have *learned* evidence E. This is, in the first place, something different from the conditional probability degree of belief in H, given E. Again, this distinction is supported by recent psychological evidence that points to differences between both modes of reasoning (Zhao et al., 2012). In that experiment, the authors found a difference between participants who learned evidence E (e.g., by observing relative frequencies) and participants who had to suppose that E occurred, in terms of the probability estimates which they submitted after the learning/supposing took place.

This finding suggests that the learning of evidence E, as it is relevant for posterior degrees of belief, should be modeled differently from supposing E (see also de Finetti, 1972; Skyrms, 1987). The obvious option is to state that the posterior probability of H after learning E should be described by the probability $p^E(H|\mathcal{M})$ which represents the agent's rational degree of belief after learning E and should be numerically identical to (but not be counterfactually interpreted as) the conditional probability $p(H|E, \mathcal{M})$. Now, it is distinctive of Bayesian reasoning that $p^E(H|\mathcal{M}) = p(H|E, \mathcal{M})$: this is just the requirement of Bayesian Conditionalization. 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 $p(H|E, \mathcal{M})$. That is, to the extent that $p(H|E, \mathcal{M})$ is constrained by the conditional degrees of belief $p(H|\mathcal{M})$, $p(E|\mathcal{M})$ and $p(E|H, \mathcal{M})$ that we happen to have, also $p^E(H|\mathcal{M})$ will be constrained. This observation rescues the normative pull of basing decisions and actions on posterior degrees of belief.

Bayesian Conditionalization relates posterior degrees of belief (that emerge from learning E) to proper conditional degrees of belief (that emerge from supposing E) in a similar way as Ratio Analysis does for conditional and unconditional probabilities. It is a *numerical constraint* that allows for the calculation of the one in terms of the other. However, it is not a *definition* or an exhaustive philosophical analysis. In both cases—posteriors vs. conditional degrees of belief, conditional vs. unconditional probabilities—the two concepts remain conceptually distinct, and the existence of a close mathematical relation between them should not seduce us to try to reduce one concept to the other.

Finally, the learning/supposing distinction also allows for a better assessment of the Problem of Old Evidence and its proposed solutions. This problem deals with the question how previously known ("old") evidence E can confirm a new theory H that manages to explain E, while other theories have been struggling with E. This pattern of reasoning is frequently found in science (e.g., Glymour, 1980), but the standard Bayesian account fails to retrieve it because the actual degree of belief in E is equal to unity. Hence, $p^E(H) = p(H|E) = p(H)$.

Most solutions of the Problem of Old Evidence make use of an argument of the form $p(E|H, \mathcal{M}) \gg p(E|\neg H, \mathcal{M})$ in the suppositional sense that we have elaborated above (e.g., Garber, 1983; Howson, 1984; Earman, 1992; Sprenger, 2015; Fitelson and Hartmann, 2016). Given H, E is way more expected than given the alternatives. This implies that E confirms H (relative to $\mathcal{M}$) on a Bayesian account. However, when the counterfactual interpretation is shunned and conditional degrees of belief are calculated according to Ratio Analysis, after learning E, these values are equal to unity: $p^E(E|H) = p^E(H \wedge E)/p^E(H) = p^E(H)/p^E(H) = 1$, and $p^E(E|\neg H) = p^E(\neg H \wedge E)/p^E(\neg H) = p^E(\neg H)/p^E(\neg H) = 1$.

The source of the problem is easily diagnosed: there is an equivocation regarding the relevant conditional degrees of belief. The interpretation that feeds the problem in the first place looks at the conditional degree of belief in E relative to H *and* all actually available information, including the actual occurrence of E. That is, it looks at the probability distribution $p^E(\cdot)$. Various solutions, by contrast, are phrased in terms of the model-relative, suppositional conditional degree of belief in E, that is, $p(E|\cdot, \mathcal{M})$, or in other words, the likelihood function $L(H) = \rho_H(E)$. Indeed, when we evaluate conditional degree of belief by counterfactually assuming that H is true, we get a different picture. In a model $\mathcal{M}$ where H competes with alternatives H', H", etc., it makes sense to say that E favors H over the alternatives because $p(E|H, \mathcal{M}) \gg p(E|H', \mathcal{M})$, $p(E|H, \mathcal{M}) \gg p(E|H'', \mathcal{M})$, etc. This is the intuition which all those who believe that old evidence can confirm a theory want to rescue. These degrees of belief can be read as a sort of ur-credences, conditional on $\mathcal{M}$ and H only (Howson, 1984; Sprenger, 2015). Making the distinction between $p^E(\cdot)$ and $p(\cdot|E, \mathcal{M})$ captures the different sorts of confirmation (or evidential favoring) that matter in science. Hence, our analysis of conditional degree of belief backs up technical solutions of the Problem of Old Evidence by a philosophical story why we can have non-trivial conditional degrees of belief in the first place. Similarly, it supports those Bayesians who believe that Bayes factors based on such conditional degrees of belief can be objective—or at least intersubjectively compelling—measures of evidence (Sprenger, 2016).

6 Conclusion

This paper was devoted to a defense of the claim that conditional degrees of belief are essentially counterfactual, taking up a proposal by Frank P. Ramsey. On the basis of this interpretation, it was argued that conditional degrees of belief equal the corresponding probability densities, as The Equality postulated ($p(H|E) = \rho_H(E)$). Furthermore, the counterfactual interpretation was extended to other probabilities in Bayesian inference, and the relation between learning evidence E and conditional degrees of belief that suppose E was investigated.

We can now state our main results. They may be shared by other philosophers of probability, but I am aware of no place where a majority of them is clearly articulated and defended.

1. Ratio Analysis is a mathematical constraint on conditional degree of belief, but provides no satisfactory philosophical analysis.

2. Conditional degrees of belief $p(E|H)$ should be interpreted in the counterfactual way outlined by Ramsey: we suppose that H were true and reason on this basis about the probability of E.

3. On this approach, supposing H will often uniquely determine the objective probability of E and the conditional degree of belief in E. In other words, we have justified The Equality: conditional degrees of belief equal the mathematical probability densities.

4. The Ramseyian approach explains the seemingly analytical nature of many probability statements in statistics, and it agrees with how statisticians view probabilities in inference: as objective, but hypothetical entities. Statistical inference does not need to be backed up by a particular interpretation of objective chance.

5. It rescues the normative pull of Bayesian inference, by contributing to agreement on the value of measures of evidential support such as the Bayes Factor.

6. It avoids problems of chance-credence coordination principles such as the Principal Principle or the Principle of Direct Inference.

7. Also other probabilities in Bayesian inference (e.g., priors) should be understood as conditional degrees of belief: they are conditional on assuming a general statistical model. This answers the challenge why we should ever have positive degrees of belief in a hypothesis when we know that the model is wrong.

8. Bayesian Conditionalization relates the posterior degrees of belief (learning E) to the conditional degrees of belief (supposing E). This should not be understood as a definition or conceptual reduction, but as a mathematical constraint that relates two distinct credence functions.

9. The distinction between posterior and conditional degrees of belief also contributes to a better assessment of the Problem of Old Evidence and various solution proposals.

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

References

Berger, J. and Wolpert, R. (1984). *The Likelihood Principle*. Manuscript under review., Hayward/CA.

Billingsley, P. (1995). *Probability and measure*. Manuscript under review., New York.

Birnbaum, A. (1962). On the Foundations of Statistical Inference. *Journal of the American Statistical Association*, 57(298):269–306.

Blackwell, D. and Dubins, L. (1962). Merging of Opinions with Increasing Information. *The Annals of Mathematical Statistics*, 33(3):882–886.

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*. Manuscript under review., New York.

de Finetti, B. (2008). *Philosophical Lectures on Probability*. Manuscript under review., Dordrecht, Holland.

Douven, I. (2016). *The Epistemology of Indicative Conditionals*. Manuscript under review., Cambridge.

Earman, J. (1992). *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Manuscript under review., Cambridge, Mass.

Easwaran, K. (2011). The Varieties of Conditional Probability. In Bandyopadhyay, P. S. and Forster, M. R., editors, *Handbook of the Philosophy of Statistics*, pages 137–148. Manuscript under review.

Fisher, R. (1956). *Statistical methods and scientific inference.* Manuscript under review., 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*.

Glymour, C. (1980). *Theory and Evidence*. Manuscript under review., Princeton.

Gyenis, Z., Hofer-Szabó, G., and Rédei, M. (2016). Conditioning using conditional expectations: The Borel-Kolmogorov Paradox.

Hacking, I. (1965). *Logic of statistical inference*. Manuscript under review., Cambridge.

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. and Urbach, P. (2006). *Scientific Reasoning: The Bayesian Approach*. Manuscript under review., La Salle, IL, 3rd edition.

Jackson, F. (1991). *Conditionals*. Manuscript under review., Oxford.

Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90:773–795.

Kolmogorov, A. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Manuscript under review., Berlin.

Kyburg, H. E. (1974). *The Logical Foundations of Statistical Inference*. Manuscript under review., Dordrecht.

Levi, I. (1977). Direct Inference. *Journal of Philosophy*, 74(1):5–29.

Lewis, D. (1980). A Subjectivist's Guide to Objective Chance. In Jeffrey, R. C., editor, *Studies in Inductive Logic and Probability*, pages 263–293. Manuscript under review., Berkeley.

Lewis, D. (1994). Humean Supervenience Debugged. *Mind*, 103:473–490.

Myrvold, W. C. (2015). You Can\textquoteright{}T Always Get What You Want Some Considerations Regarding Conditional Probabilities. *Erkenntnis*, 80(3):573–603.

Popper, K. R. (2002). *The Logic of Scientific Discovery*. Manuscript under review., London, reprint edition.

Ramsey, F. P. (1926). Truth and Probability. In Mellor, D. H., editor, *Philosophical Papers*, pages 52–94. Manuscript under review., Cambridge.

Reichenbach, H. (1949). *The Theory of Probability*. Manuscript under review.

Renyi, A. (1970). *Foundations of Probability*. Manuscript under review., San Francisco.

Rosenthal, J. (2004). *Wahrscheinlichkeiten als Tendenzen*. Manuscript under review., Paderborn.

Savage, L. J. (1972). *The Foundations of Statistics*. Manuscript under review., New York, 2nd edition.

Skyrms, B. (1987). Updating, Supposing, and Maxent. *Theory and Decision*, 22(3):225–246.

Skyrms, B. (2000). *Choice & Chance*. Manuscript under review., Belmont, CA, 4th edition.

Sober, E. (2008). *Evidence and Evolution: The Logic Behind the Science*. Manuscript under review., Cambridge.

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. Manuscript under review.

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*. Manuscript under review., 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*. Manuscript under review., 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. Manuscript under review., London.

Williamson, J. (2010). *In Defence of Objective Bayesianism*. Manuscript under review., 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.