Conditional Degree of Belief

Jan Sprenger*

July 25, 2016

Abstract

This paper investigates the semantics and epistemology of conditional degree of belief. It defends the thesis that conditional degrees of belief should be interpreted counterfactually. This interpretation is the only one that explains the guiding role of conditional degree of belief in probabilistic reasoning and the normative pull of Bayesian inference. Then, I explore the implications of this thesis for the meaning of Bayes' Theorem, chance-credence coordination and the Problem of Old Evidence.

*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

1 Introduction

The normative force of Bayesian inference is crucially affected by the way one understands conditional degree of belief. To see this, note that Bayesians represent subjective degrees of belief by a particular mathematical structure—probabilities—and change them by the principle of Conditionalization: the degree of belief in a hypothesis H after learning evidence E is equal to the conditional probability of H given E. In other words, $p^{\mathrm{E}}(\mathrm{H}) = p(\mathrm{E}|\mathrm{H})$.

The latter probability is, in turn, crucially affected by the probability of E given H and its negation $\neg$H, as evident from Bayes' famous theorem:

$$
\begin{aligned}
p(\mathrm{H}|\mathrm{E}) &= p(\mathrm{H})\,\frac{p(\mathrm{E}|\mathrm{H})}{p(\mathrm{E})} && \text{(Bayes' Theorem)} \\
&= \left(1 + \frac{1 - p(\mathrm{H})}{p(\mathrm{H})}\,\frac{p(\mathrm{E}|\neg\mathrm{H})}{p(\mathrm{E}|\mathrm{H})}\right)^{-1}
\end{aligned}
$$

The theorem demonstrates that the conditional degree of belief in H given E, and thereby also the posterior probability of H, $p^{\mathrm{E}}(\mathrm{H})$, is a function of the prior degree of belief in H and the conditional degrees of belief in E given H and $\neg$H.

Bayesians use the posterior probability of H as a basis for assessing theories and making practical decisions (Jeffrey, 1971; Savage, 1972; Howson and Urbach, 2006). Therefore, the normative force of Bayesian inference relies to a large extent on how the conditional degrees of belief in evidence E given H and $\neg$H are constrained. After all, (subjective) Bayesians typically allow for a wide range of rational prior degrees of belief in H. If we want to defend Bayesian inference against the charge of total arbitrariness, the conditional degrees of belief $p(\mathrm{E}|\mathrm{H})$ and $p(\mathrm{E}|\neg\mathrm{H})$ cannot just take any value. This motivates a detailed analysis of conditional degrees of belief, and how they are constrained in Bayesian inference.

Two further observations shall back up this approach. First, agreement on $p(\mathrm{E}|\mathrm{H})$ and $p(\mathrm{E}|\neg\mathrm{H})$ is required to support 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 collect observations. When evidence $\mathrm{E}_1, \ldots, \mathrm{E}_\mathrm{n}$ accumulates, the ratio of $p(\mathrm{E}_1, \ldots, \mathrm{E}_\mathrm{n}|\neg\mathrm{H})$ and $p(\mathrm{E}_1, \ldots, \mathrm{E}_\mathrm{n}|\mathrm{H})$ will, in the limit $n \to \infty$, eventually dominate the ratio of prior probabilities in Bayes' Theorem. Therefore, $p(\mathrm{H}|\mathrm{E})$ will converge either to one or to zero, dependent on whether H or $\neg$H is true. However, if the conditional degrees of belief $p(\mathrm{E}_\mathrm{n}|\mathrm{H})$ and $p(\mathrm{E}_\mathrm{n}|\neg\mathrm{H})$ were to vary, agents would also reach different posterior probabilities. Consensus on the true hypothesis would be impossible.

Second, think of Bayesian statistical inference. A statistical model consists of a sample space $\mathcal{S}$ together with a set of probability distributions over $\mathcal{S}$ (e.g. Cox and Hinkley, 1974; Bernardo and Smith, 1994; McCullagh, 2002). If the model is parametric—as most scientific models are—then each θ from the parameter set Θ is assigned a particular probability distribution which describes the probability of events E by means of a *probability density function* $\rho_\theta(\mathrm{E})$.

For example, suppose that we want to model a repeated coin toss. The hypotheses $(\mathrm{H}_\theta, \theta \in [0,1])$ describe the chance of the coin to come up heads on any individual toss. When we toss the coin N times, our sample space is $\mathcal{S} = \{0,1\}^N$. Under the assumption that the tosses are independent and identically distributed (henceforth, i.i.d.), we can describe the probability of observation E_k (=k heads and $N-k$ tails) by the Binomial probability distribution and the corresponding density function $\rho_{\mathrm{H}_\theta}(\mathrm{E}_k) = \binom{N}{k}\theta^k(1-\theta)^{N-k}$. Bayesians routinely plug in such probability densities for the values of $p(\mathrm{E_k}|\mathrm{H}_\theta)$ that figure in Bayes' Theorem. These strong constraints on conditional degree of belief often lead to approximate agreement on posterior probabilities, as long as the priors are not too extreme.

Third, when conditional degrees of belief equal the corresponding values of probability density functions, we can straightforwardly 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 conditional degrees of belief are unconstrained, there may be no fact of the matter as to whether E favors H_0 over H_1, or vice versa. For the application of Bayesian inference in the sciences where such judgments of evidential favoring are urgently required, 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.

Hence, explaining why conditional degrees of belief in E given H and $\neg$H are rationally constrained is vital for those Bayesians who would like to defend the objectivity and intersubjective appeal of Bayesian inference in science (e.g., Bernardo, 2012; Sprenger, 2016). Moreover, it is a vital issue for all Bayesian reasoning about belief, evidence and rational decisions. This brings us to our main question:

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

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

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 these principles apply to Bayesian inference in science? Only partially so. A (statistical) hypothesis H describes the probability of possible events E by the density function $\rho_H(\mathrm{E})$, which satisfies the axioms of probability. So ρ_H is a probability function. Moreover, ρ_H does not depend on subjective epistemic attitudes, but just on the definition of the statistical model. So it is objective. But not all objective probabilities are ontic probabilities, that is, physical chances that could figure in PDI or PP (Rosenthal, 2004). Sentences that feed typical applications of PP and PDI, such as "the chance of tossing two heads in two tosses of this coin is 1/4", make an empirical statement and do not refer to a statistical model. Their truth conditions depend on the properties of the (physical) coin and the tosses. Probability density functions, on the other hand, do not describe properties of an object or an experimental setup. The truth value of sentences such as $\rho_H(\mathrm{E}) = 1/4$ is entirely internal to the statistical model: if H denotes the hypothesis that the coin is fair and E the observation of two tosses, then it is part of the *meaning* of H that the probability of E given H is 1/4 (see also Sprenger, 2010). That is, the sentence

> If a fair coin is tossed repeatedly, then the chance of two heads in two i.i.d. tosses is 1/4.

has no empirical content—it has a distinctly analytical flavor. It does not refer to chances which are instantiated anywhere in the real world; in fact, there need not exist any fair coins for this sentence to be true. Many scientific models are of this idealized type and do not have any real-world implementation (Frigg and Hartmann, 2012). The Principle of Direct Inference and the Principal Principle, by contrast, apply to physical chances: there has to be an event whose actual, objective chance is 1/4. PDI and PP 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—and at the same time, to explain which role chance-credence coordination plays in Bayesian reasoning.

Hence, our paper joins the efforts by philosophers of science to clarify the nature of objective probability in scientific reasoning. Hoefer (2007) articulated this 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 probability density functions in informing and constraining conditional degree of belief has barely been investigated. At this point, I would like to stress that statistical inference serves as a neat case study for studying the semantics and epistemology of conditional degree of belief, but that my theses to more general cases of Bayesian inference, too.

The rest of the paper is structured as follows. Section 2 examines a reductive analysis of conditional degree of belief: the ratio analysis of conditional probability. I argue that it fails to account for The Equality and to explain the practice of Bayesian reasoning. Section 3 develops a semantics of conditional degrees of belief: they should be interpreted counterfactually, similar to what Frank P. Ramsey proposed almost a century ago. We then elaborate on how this proposal justifies The Equality, that is, the agreement between conditional degrees of belief and probability densities, and we explain the role of chance-credence coordination in deriving The Equality. Section 4 explores the implications of our proposal for Bayesian inference and extends the counterfactual interpretation of conditional degree of belief to prior and posterior probabilities. Thereby we explain why Bayesians can assign positive degrees of belief to a hypothesis which they know to be wrong. Section 5 discusses the roles of learning and supposing in Bayesian inference in the light of our proposal. 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).

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

By applying a Dutch Book argument, we can transfer this reductive analysis of conditional probability to conditional degree of belief (e.g., Easwaran, 2011a). It can be shown that the conditional degree of belief in E given H, operationalized by a bet on E that is called off if $\neg$H, must be equal to the ratio of the unconditional degrees of belief in $\mathrm{E} \wedge \mathrm{H}$ and H if a Dutch book is to be avoided. Therefore one might conjecture that our search for a justification of The Equality is a *Scheinproblem*: conditional degrees of belief are determined by the coherence constraints which Ratio Analysis imposes on them, and this explains why they have to satisfy The Equality.

I do not find this view convincing, but for reasons that differ from the usual ones. Here is the standard objection to Ratio Analysis. 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 $\theta \in [0, 1]$. Usually, each single point in $[0, 1]$, which corresponds to a particular hypothesis about the bias of the coin, will have probability zero. In this case, 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 two 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 E given 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 (E), given that it lies on the equator (H)?". 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, $p(\mathrm{H}) = 0$. Notably, even the well-known probability theory textbook by Billingsley (1995, 427) writes that "the conditional probability of [E] with respect to [H, ...] is defined [by Ratio Analysis] unless $p(\mathrm{H})$ vanishes, in which case it is not defined at all" (notation adapted). In other words, the wide variety of cases where we have conditional degrees of belief is not adequately captured by Ratio Analysis. The seminal article by Alan Hájek (2003)—"What Conditional Probability Could Not Be"—discusses these problems in detail.

However, the probability zero objection is not necessarily a knock-down argument. Howson and Urbach (e.g., 2006, 37) propose the following fix. If H: $\theta = \theta_0$ for some parameter value $\theta_0 \in \Theta \subset \mathbb{R}$ and $p(\mathrm{H}) = 0$, then define $\mathrm{H}_\varepsilon : \theta \in (\theta_0 - \varepsilon, \theta_0 + \varepsilon)$. Then we define

$$p(\mathrm{E}|\mathrm{H}) := \lim_{\varepsilon \to 0} \frac{p(\mathrm{E} \wedge \mathrm{H}_\varepsilon)}{p(\mathrm{H}_\varepsilon)}$$

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 $p(E|H_\theta)$.

Hájek mentions another refinement of Ratio Analysis that may help to deflect the probability zero objection. It 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. Easwaran (2011b) 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 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). What I wanted to point out is that the probability zero objection is a promising, but not necessarily conclusive attack on Ratio Analysis as a definition of conditional probability. My own line of attack is therefore different: even if these technical problems were solved, Ratio Analysis would fail to give an adequate explanation of how we reason with conditional degrees of belief, even in those cases where $p(\mathrm{H}) > 0$. It 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 degrees of belief: they are just 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. If our degree of belief in the occurrence of heads on two tosses of a fair coin is elicited, nobody will first calculate her unconditional degree of belief in an occurrence of two heads, and her degree of belief in the conjunction of two heads and the fairness hypothesis. It is cognitively just too demanding to elicit $p(\mathrm{E} \wedge \mathrm{H})$ and $p(\mathrm{H})$ and to calculate their ratio. Instead, we directly assess how probable E is, given H. This is more than a purely phenomenological argument: recent psychological evidence demonstrates 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 $p(\mathrm{E}|\mathrm{H})$ 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 two tosses is $1/4$ *given that the coin is fair*. On Ratio Analysis, this feature of conditional degree of belief drops out of the picture. $p(\mathrm{E}|\mathrm{H})$ is constrained only via constraints on $p(\mathrm{E} \wedge \mathrm{H})$ and $p(\mathrm{H})$. But even if we suspend judgment on $p(\mathrm{E} \wedge \mathrm{H})$ and $p(\mathrm{H})$, there are still constraints on $p(\mathrm{E}|\mathrm{H})$. Ratio Analysis therefore misses an important aspect of conditional degree of belief.

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. Unsurprisingly, Fitelson and Hájek (2016) advocate replacing Kolmogorov's axioms by the Popper-Rényi axioms (Rényi, 1970; Popper, 2002) where conditional probability is take as primitive. This emphasizes the need for a proper semantics of conditional degree of belief, which I will provide in the next section.

3 The Counterfactual Analysis

Between the lines, the previous sections have 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 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 be a verifiable event if $p(\mathrm{E}|\mathrm{H})$ is 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.

I would like to stay neutral on all issues concerning conditionals (e.g., Douven, 2016) 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, $p(\mathrm{E}|\mathrm{H})$ is the rational degree of belief in E if we supposed that H were true.

In other words, this analysis of conditional degree of belief asks us to imagine a possible world ω_H, created by supposing H. ω_H is genuinely chancy, regardless of whether the actual world is: the possible events are elements of the sample space $\mathcal{S}$, and their probability is described by ρ_H. This is the essence of supposing a statistical model and a particular hypothesis within that model. It is the *meaning* of $\rho_H(E)$ to express the chance of E if H were the case. The (physical) chance of two heads in two tosses in ω_H is $1/4$. 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 possible events. Therefore, the counterfactual analysis of conditional degree of belief directly yields The Equality.

Incidentally, this interpretation of conditional degree of belief differs from Ramsey's in a crucial nuance: where Ramsey suggested that H is *added* to existing background knowledge, I propose that H *overrules* conflicting knowledge about the real world, leading to a genuinely counterfactual interpretation of conditional degree of belief. Even if we know that a given coin is not perfectly fair and that the tosses are not i.i.d., the conditional degree of belief in the occurrence of k heads in N tosses, given a statistical model of the coin toss, is derived from supposing that the modeling assumptions are valid. Indeed, reasoning with scientific models is typically counterfactual and involves substantial abstraction steps (e.g., Batterman, 2002; Weisberg, 2007; Frigg and Hartmann, 2012). Statistical models are no exception.

It is important to understand the role of the Principle of Direct Inference (PDI) and the Principal Principle (PP) in this picture. 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 do not deal with 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 $\rho_H(E)$. In other words, we do not apply PDI/PP in the actual world $\omega_@$, but in the counterfactual world ω_H described by H, 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 counterfactual 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.

I would like to add that complications induced by inadmissible information (e.g., Lewis, 1980; Hoefer, 2007) do not occur in our case because the worlds ω_H are so simple and well-behaved. The events in ω_H are elements of the sample space $\mathcal{S}$ and H assigns a definite and unambiguous probability to them. More on the interaction of H with past observations shall be said in Section 5. For the moment, we can explain why chance-credence coordination is so important for probabilistic reasoning without committing ourselves to the existence or nature of physical chance in the actual world. No scientist needs to reason about the nature of physical chance when she uses a statistical model to inform her credences. It is a distinct strength of this analysis of conditional degree of belief that it is compatible with any account of physical chance.

Let us wrap up the essence of my proposal. Bayesian inference requires two things for coordinating conditional degrees of belief with probability densities: First, the counterfactual interpretation of conditional degree of belief which supposes that H is true—even if we actually know that it is false. Second, in these counterfactual scenarios, degrees of beliefs should be coordinated with known objective chances (Lewis, 1980). To repeat, we are talking about chance-credence coordination in hypothetical worlds where the space of possible events (=the sampling space) is determined by the statistical model, 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 density functions inform our *hypothetical* degrees of belief, not our actual degrees of belief.

The proposed interpretation matches the thoughts of the great (non-Bayesian) statistician Ronald A. Fisher on the nature of conditional probability in scientific 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 (Gallistel, 2009). 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. 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. We now proceed to its wider implications for Bayesian inference.

4 Implications I: The Model-Relativity of Bayesian Inference

An obvious objection to the picture sketched in the previous section concerns the demarcation of the statistical hypothesis H (e.g., that the parameter θ takes a specific value) and the other assumptions which are part of the statistical model $\mathcal{M}$. Consider the case of tossing a coin. When we evaluate $p(\mathrm{E}|\mathrm{H})$ with H = "the coin is fair", we already assume that the individual tosses of the coin are independent and identically distributed. However, this assumption is not part of H itself: H just describes the tendency of the coin on any particular toss. If we contrast H to some alternative H′, we notice that the differences between them are typically expressed in terms of parameter values, such as H: $\theta = 1/2$ versus H′: $\theta = 2/3$, H″: $\theta > 1/2$, etc. Crucial assumptions on the experimental setup, such as independence and identical distribution of the coin tosses, do not enter the particular hypothesis we are testing. Does this jeopardize our analysis of conditional degree of belief? No. These assumptions are already part of general model in which we compare H to H′. In other words, there are two layers in the modeling process. First, there is the general statistical model $\mathcal{M} = (\mathcal{S}; \mathcal{P})$ where $\mathcal{S}$ denotes the sampling space, that is, the set of possible observations in an experiment, and $\mathcal{P}$ denotes the set of probability distributions over $\mathcal{S}$. Second, there is the layer of the individual statistical hypotheses which are elements of $\mathcal{P}$. Considering only a narrow set of hypotheses (e.g., the set of Binomial distributions $B(N, \theta)$) makes an implicit assumption on the experimental setup (e.g., that the tosses are i.i.d.): the hypotheses do not differ from each other in how they describe several aspects of the experiment.

This implies that the conditional degree of belief $p(\mathrm{E}|\mathrm{H})$ 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 $p(\mathrm{H})$: Bayesian inference regarding particular parameter values is relative to a model $\mathcal{M}$ into which all hypotheses are embedded (e.g., $\mathcal{M} = (\{0,1\}^N; B(N,\theta), \theta \in [0,1])$), and degrees of belief are distributed only over elements of $\mathcal{M}$. In particular, also the prior and posterior degrees of belief, $p(\mathrm{H})$ and $p(\mathrm{H}|\mathrm{E})$, are essentially 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 hypothesis H, $p(\mathrm{H})$, 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 hypothesis. It would be even more silly to bet on the truth of any particular model, as operational interpretations of subjective probability demand. Also the sample 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 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 $p(\mathrm{H})$. 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 $p(\mathrm{H})$, we talk about $p(\mathrm{H}|\mathcal{M})$. This move makes the entire Bayesian inference relative to $\mathcal{M}$. Similarly, we replace the marginal likelihood $p(\mathrm{E})$ by $p(\mathrm{E}|\mathcal{M})$ and interpret it counterfactually, in agreement with the Law of Total Probability. $p(\mathrm{E}|\mathcal{M})$ is the weighted average of the conditional probabilities of E, and thus our subjective expectation that E occurs if $\mathcal{M}$ were the case.

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

However, we notice that not all conditional degrees of belief are of the same kind. There is a relevant difference between $p(\mathrm{E}|\mathrm{H}, \mathcal{M})$ on the one hand and $p(\mathrm{H}|\mathcal{M})$ 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 $\rho_{\mathcal{M},H}(\mathrm{E})$. However, supposing $\mathcal{M}$ does not yield a uniquely rational value for $p(\mathrm{H}|\mathcal{M})$. There is no objective chance of H in the hypothetical world $\omega_{\mathcal{M}}$, nor a corresponding uniquely determined probability density. Does this mean that the counterfactual analysis of conditional degree of belief only applies to those cases where a probability density function fixes the rational degree of belief?

To my mind, the answer is no. The fact that $p(\mathrm{H}|\mathcal{M})$ is not uniquely determined does not mean that it cannot be interpreted counterfactually. People may reason differently about what the degree of belief in H, given $\mathcal{M}$, should be. But it is part and parcel of subjective Bayesianism that these choices are essentially unconstrained. Some choices may lead to a better match between the statistical model and the target system (e.g., the repeated coin toss) than others. But this is a matter of which priors are better calibrated with our knowledge about the external world, and not a question that can be answered from 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. Hence, Bayesian inference makes sense even if all models are known to be wrong, as long as some are illuminating and useful (Box, 1976). In this picture, it is evident that Bayesian inference is just another way of model-based reasoning in science.

The same remarks apply to the posterior probability $p(\mathrm{H}|\mathrm{E}, \mathcal{M})$. By itself, supposing E and $\mathcal{M}$ 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, since the world $\omega_{\mathcal{M},E}$ does not determine an objective chance of H. This points to a divergence in our analysis of conditional degree of belief: for probabilities of the type $p(\mathrm{E}|\mathrm{H}, \mathcal{M})$, supposing H and $\mathcal{M}$ constraints the conditional degrees of belief in an objective way, but supposing E and $\mathcal{M}$ does not deliver such a result for $p(\mathrm{H}|\mathcal{M})$ or $p(\mathrm{H}|\mathrm{E}, \mathcal{M})$. How can eventual agreement on posterior probabilities be rescued if this diagnosis is correct? In a very simple way: $p(\mathrm{H}|\mathrm{E}, \mathcal{M})$ is constrained to the extent that $p(\mathrm{E}|\mathrm{H}, \mathcal{M})$ $p(\mathrm{E}|\mathcal{M})$ and $p(H, \mathcal{M})$ are, by means of Bayes' Theorem. Assume that we have a sharp degree of belief $p(\mathrm{H}|\mathcal{M})$, as a matter of psychological fact, or by calibrating $p(\mathrm{H}|\mathcal{M})$ with our background knowledge. Assume further that we have a sharp degree of belief $p(\mathrm{E}|\pm\mathrm{H}, \mathcal{M})$, by supposing H and $\neg$H, as explained in Section 3. Then our degrees of belief in $p(\mathrm{E}|\mathcal{M})$ and $p(\mathrm{H}|\mathrm{E}, \mathcal{M})$ are fixed as well because $p(\cdot|\mathcal{M})$ is a probability function and therefore satisfies Bayes' Theorem:

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

If our conditional degrees of belief do not satisfy this equality, we will also violate Ratio Analysis, run into a Dutch Book and violate the Bayesian norms of rationality. The normative pull for constraining $p(\mathrm{H}|\mathrm{E}, \mathcal{M})$ does not emerge directly from the counterfactual interpretation of conditional degrees of belief (=supposing E and $\mathcal{M}$), but from the requirement of probabilistic coherence.

This reading of Bayesian inference is very close to Ramsey's and de Finetti's subjective Bayesianism: 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 subjective Bayesianism should not be confused with an "anything goes" attitude: rationality constraints for a subjective Bayesian can be substantial.

This brings us to the role of Bayes' Theorem. First, a misunderstanding needs to be avoided. 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 in the next section.) It is a synchronic, not a diachronic constraint, and Bayesian Conditionalization has not been invoked in stating the above equality.

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

Third, on the counterfactual interpretation, we can regard Bayes' Theorem as a coordinating principle for various probability function that describe our conditional degrees of belief. Rather than a theorem of mathematics that applies to a single probability function, it states how different functions representing conditional degree of belief ($p(\cdot|\mathcal{M})$, $p(\cdot|\mathrm{H}, \mathcal{M})$ and $p(\cdot|\mathrm{E}, \mathcal{M})$) should coordinate in order to avoid a Dutch book. This way of looking at the theorem elucidates why it has philosophical significance and why it is something else than a simple mathematical result.

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 in 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 ($p(\perp) = 0$; $p(A) + p(\neg A) = 1$; $p(\vee A_i) = \sum p(A_i)$ 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-Rényi axioms (Rényi, 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.

So far, we have discussed the statics of conditional degrees of belief. But what about the dynamics? The next section tries to give answers.

5 Implications II: Learning vs. Supposing

Learning and supposing are two major elements of Bayesian inference and we will now study their interaction in greater detail. To begin with, experiments in cognitive psychology have shown that they are really different modes of reasoning. In a recent study, Zhao et al. (2012) 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.

On a theoretical level, we have postulated that supposing informs conditional degree of belief. Learning, on the other hand, happens by means of the principle of Conditionalization which governs the dynamics of Bayesian reasoning. Essentially, Conditionalization relates learning E to supposing E: for a hypothesis H and an observation E, $p^{\mathrm{E}}(\mathrm{H}) = p(\mathrm{H}|\mathrm{E})$. That is, our degree of belief in H after learning E should be equal to the degree of belief that we would have in H if we supposed E. There are different ways how Conditionalization can be justified and attacked (e.g., Teller, 1973; van Fraassen, 1989; Easwaran, 2011a). Here, we are just interested in how Conditionalization fits into the counterfactual interpretation and model-relativity of conditional degree of belief. In particular, how does learning the observation E^* affect the conditional degree of belief $p^{\mathrm{E}^*}(\mathrm{E}|\mathrm{H})$?

In a nutshell, the answer is that we evaluate this probability relative to the counterfactual world ω_H, taking into account the occurence of E^* within that world.

This gives rise to several possible cases. First, E^* could be impossible given H, such as in the case where H asserts that a coin always comes up heads, and E^* denotes an observation of tails. In this case, the entire scenario is inconsistent and the conditional probability $p^{\mathrm{E}^*}(\mathrm{E}|\mathrm{H}, \mathcal{M})$ would not be well-defined.

Second, the case where E^* and H do not contradict each other. Sometimes $p^{\mathrm{E}^*}(\mathrm{E}|\mathrm{H}, \mathcal{M})$ will then be equal to $p(\mathrm{E}|\mathrm{H}, \mathcal{M})$, e.g., when E and E^* denote two i.i.d. tosses of a coin and H screens off both observations from each other. In other cases, a dependency is possible. Another example are time series. Consider $\mathcal{M} : Y_t = Y_{t-1} + X_t$ with $X_t \in \{-1, 1\}$ and the hypothesis H: $p(X_t = 1) = 1/2$. If $\mathrm{E}^* : X_{t-1} = 3$ and E: $X_t = 4$, then we can directly calcluate that $p^{\mathrm{E}^*}(\mathrm{E}|\mathrm{H}, \mathcal{M}) = 1/2$. Statistical model, hypothesis and past evidence constrain $p^{\mathrm{E}^*}(\mathrm{E}|\mathrm{H}, \mathcal{M})$ to a unique value.

The above reading of Bayesian learning introduces an implicit distinction between propositions on a theoretical and observational level. After all, supposing $\mathcal{M}$ and H specifies the sample space of possible observations, whereas E^* corresponds to some observation within this space. While the distinction is perhaps difficult to maintain on a general philosophical level, it is sharp for most important applications of Bayesian inference (e.g., statistical hypotheses vs. elements of the sample space).

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: if E is already known then $p(\mathrm{E}) = 1$ and $p^E(H) = p(\mathrm{H}|\mathrm{E}) = p(\mathrm{H}) \cdot p(\mathrm{E}|\mathrm{H})/p(\mathrm{E}) \leq p(\mathrm{H})$.

Most solutions of the Problem of Old Evidence use an argument of the form $p(\mathrm{E}|\mathrm{H}, \mathcal{M}) \gg p(\mathrm{E}|\neg\mathrm{H}, \mathcal{M})$ with a counterfactual interpretation of these probabilities (e.g., Garber, 1983; Howson, 1984; Earman, 1992; Sprenger, 2015; Fitelson and Hartmann, 2016). Given a scientific framework $\mathcal{M}$ and a particular hypothesis H in this framework, E is way more expected under H than under the available alternatives. This counterfactual reading allows us to "forget" that E has been observed and to obtain a meaningful difference between the two conditional probabilities. Hence, E confirms H (relative to $\mathcal{M}$) on a Bayesian account. However, it is not clear how this observation bears on a situation where E is already known.

Apparently, 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 $p^E(\cdot)$. Then, all relevant 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 $p^E(E|\neg H) = p^E(\neg H \wedge E)/p^E(\neg H) = p^E(\neg H)/p^E(\neg H) = 1$.

Various solutions, by contrast, are phrased in terms of the counterfactual conditional degree of belief in E, that is, $p(\mathrm{E}|\cdot, \mathcal{M})$. 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 $p(\mathrm{E}|\mathrm{H}, \mathcal{M}) \gg p(\mathrm{E}|\mathrm{H}', \mathcal{M}), p(\mathrm{E}|\mathrm{H}'', \mathcal{M})$, 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, and our account provides the conceptual framework for expressing these credences (Howson, 1984; Sprenger, 2015).

Other solutions of the Problem of Old Evidence work with the probability function $p^E(\cdot)$ and aim to show that learning H $\vdash$ E confirms H relative to this function (Jeffrey, 1983; Earman, 1992). 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, developing 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 ($p(\mathrm{H}|\mathrm{E}) = \rho_H(E)$). 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 they are clearly articulated and defended.

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

2. Conditional degrees of belief $p(\mathrm{E}|\mathrm{H})$ 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. The Equality follows directly from this counterfactual interpretation: conditional degrees of belief in an event given a hypothesis follow the corresponding probability densities.

4. The counterfactual interpretation explains the seemingly analytical nature of many probability statements in statistics, and it agrees with how scientists view

probabilities in inference: as objective, but hypothetical entities. Scientific inference with probabilities need not be backed up by a particular interpretation of objective chance.

5. The counterfactual interpretation 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. The counterfactual interpretation explains the role of chance-credence coordination principles in scientific inference, such as the Principal Principle or the Principle of Direct Inference.

7. All probabilities in Bayesian inference are conditional degrees of belief: they are conditional on assuming a general statistical model. This approach counters the objection that we should never have positive degrees of belief in a hypothesis because we know it to be wrong.

8. Bayes' Theorem expresses an epistemic coordination principle for various functions that describe conditional degree of belief.

9. Bayesian Conditionalization relates the posterior degrees of belief (learning E) to the conditional degrees of belief (supposing E).

10. The learning-supposing distinction and the counterfactual interpretation of degrees of belief jointly lead to a better assessment of the Problem of Old Evidence and the various solution proposals.

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

References

Batterman, R. W. (2002). *The devil in the details: asymptotic reasoning in explanation, reduction, and emergence*. Oxford University Press.

Bernardo, J. M. (2012). Integrated objective Bayesian estimation and hypothesis testing. In *Bayesian Statistics 9: Proceedings of the Ninth Valencia Meeting*, pages 1–68 (with discussion). Oxford University Press, Oxford.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. Wiley, New York.

Billingsley, P. (1995). *Probability and Measure*. Wiley, New York.

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

Box, G. E. P. (1976). Science and Statistics. *Journal of the American Statistical Association*, 71:791–799.

Cox, D. and Hinkley, D. (1974). *Theoretical Statistics*. Chapman & Hall, London.

de Finetti, B. (1972). *Probability, Induction and Statistics: the Art of Guessing*. Wiley, New York.

de Finetti, B. (2008). *Philosophical Lectures on Probability*. Springer, Berlin.

Douven, I. (2016). *The Epistemology of Indicative Conditionals*. Cambridge University Press, Cambridge.

Earman, J. (1992). *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. The MIT Press, Cambridge/MA.

Easwaran, K. (2011a). Bayesianism I: Introduction and Arguments in Favor. *Philosophy Compass*, 6:312–320.

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

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*, 47:495–548.

Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review*, 116:439–453.

Garber, D. (1983). Old Evidence and Logical Omniscience in Bayesian Confirmation Theory. In Earman, J., editor, *Testing scientific theories*, pages 99–132. University of Minnesota Press, Minneapolis.

Glymour, C. (1980). *Theory and Evidence*. Princeton University Press, Princeton.

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

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*, 116:549–596.

Howson, C. (1984). Bayesianism and support by novel facts. *British Journal for the Philosophy of Science*, 35:245–251.

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.

Jeffrey, R. (1971). *The Logic of Decision*. University of Chicago Press.

Jeffrey, R. C. (1983). Bayesianism with a human face. In Earman, J., editor, *Testing scientific theories*, pages 133–156. University of Minnesota Press, Minneapolis.

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*. Springer, Berlin.

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

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

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

McCullagh, P. (2002). What is a statistical model? *Annals of Statistics*, 30:1225–1310.

Myrvold, W. C. (2015). You Can't Always Get What You Want: Some Considerations Regarding Conditional Probabilities. *Erkenntnis*, 80:573–603.

Popper, K. R. (2002). *The Logic of Scientific Discovery*. Routledge, London. Reprint of the revised English 1959 edition. Originally published in German in 1934 as "Logik der Forschung".

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

Reichenbach, H. (1949). *The Theory of Probability*. University of California Press, Berkeley/CA.

Rényi, A. (1970). *Foundations of Probability*. Holden-Day, San Francisco.

Rosenthal, J. (2004). *Wahrscheinlichkeiten als Tendenzen*. mentis, Paderborn.

Savage, L. J. (1972). *The Foundations of Statistics*. Wiley, 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, Berlin.

Sprenger, J. (2015). A Novel Solution of the Problem of Old Evidence. *Philosophy of Science*, 82:383–401.

Sprenger, J. (2016). The Objectivity of Subjective Bayesian Inference. Manuscript under review.

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*, 5:269–286.

Suarez, M. (2011a). *Probabilities, Causes and Propensities in Physics*. Springer, New York.

Suarez, M. (2011b). Propensities and Pragmatism. *Journal of Philosophy*, 100:61–92.

Teller, P. (1973). Conditionalization and Observation. *Synthese*, 26:218–258.

van Fraassen, B. C. (1989). *Laws and Symmetry*. Oxford University Press, New York.

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*, 113:26–36.