# Abducted by Bayesians?

Jan-Willem Romeijn
Department of Philosophy
University of Groningen[*]
j.w.romeijn@rug.nl

**Abstract**

This paper discusses the role of theoretical notions in making predictions and evaluating statistical models. The core idea of the paper is that such theoretical notions can be spelled out in terms of priors over statistical models, and that such priors can themselves be assigned probabilities. The discussion substantiates the claim that the use of theoretical notions may offer specific empirical advantages. Moreover, I argue that this use of theoretical notions explicates a particular kind of abductive inference. The paper thus contributes to the discussion over Bayesian models of abductive inference.

## 1 Introduction

In this section I introduce theoretical notions in a statistical context. Against this background I specify the two central claims of this paper.

### 1.1 Theoretical notions and empirical content

In what follows I adopt the view that the empirical content of a statistical hypothesis $H$ is given by its likelihoods, that is, by the probabilities of data $E$ conditional on the hypothesis, written as $P(E|H)$ (cf. Douven [2008]). If two hypotheses $H$ and $H^{\star}$ have identical likelihoods, that is,

$$P(E|H) = P(E|H^{\star})$$

[*] The author is simultaneously a research fellow at the Philosophy Department of the University of Johannesburg.

for all possible observations $E$, then they have the same empirical content.[1] We will say that a distinction between hypotheses is based on a theoretical notion, or theoretical for short, if it distinguishes two hypotheses $H$ and $H^{\star}$ while these hypotheses have identical likelihood functions. Furthermore, a statistical model is defined to be a collection of hypotheses, $\mathcal{H} = \{H_1, H_2, \ldots, H_n\}$. We will say that a distinction between two models $\mathcal{H}$ and $\mathcal{H}^{\star}$ is theoretical if the hypotheses in the model have pairwise identical likelihood functions, $P(E|H_j) = P(E|H^{\star}_j)$ for $0 < j \leq n$ and for all $E$.

A central point in this paper is that models whose distinction is theoretical may still differ in empirical content, because of the priors we define over them. We will look at models $\mathcal{H}$ and $\mathcal{H}^{\star}$ that differ theoretically in the sense specified above, but that are associated with different stories concerning the data generating system. Such stories motivate different priors over the models in question, and these priors again lead to a different empirical content for the two models. The data may be used to choose between the models in virtue of their association with different priors.

This approach to comparing statistical models has been around for a while, and indeed can be traced back to Gaifman [1985].[2] It is perfectly coherent to assign probabilities to probability assignments, as is routinely done in Bayesian statistics, and it is also coherent to add more layers of probabilistic analysis, assigning probabilities to the probability assignments over the groud level assignments, and so on. More recently, this idea has taken root in statistics, more precisely in hierarchical Bayesian modeling (cf. Gelman *at al* [2004], Gelman and Hill [2007]). This approach compares models on their marginal likelihoods, in which the prior over the model is an explicit component. Henderson *et al* [2010] discuss some philosophical applications of hierarchical modeling.

[1] If we assume the likelihood principle, according to which all evidence pertaining to a hypothesis is mediated by the likelihoods of that hypothesis, then the two hypotheses cannot be distinguished by any evidence. See Royall [2000].

[2] Considering the infuence of Prof. Gaifman's work on the philosophy of science, and on the philosophy of statistics in particular, the connection of this paper to his work is a rather weak one. I regret that the philosophical appraisal of Gaifman and Snir's seminal paper on rich languages, which was planned for this special issue, is still under construction.

### 1.2 Central claims of this paper

The upshot of this paper is that theoretical notions can play an active role in statistical inference, and that in particular cases we can tell apart theoretically distinct models by empirical means. Higher-order probabilities play a crucial role in capturing these theoretical notions and in making them empirical. These two claims answer two critical discussions on theoretical notions in science. The first of these is that theoretical notions can be understood as methodological tools, and should not be discarded as as superfluous and non-empirical. This claim can be viewed a response to the theoretician's dilemma discussed in Hempel [1958]. He argues that, if the aims of science are indeed empirical, there is no role for theoretical notions, since we can purge a scientific theory of such notions without losing any of its empirical content. Against this, I argue that certain uses of theoretical notions lead to more efficient inferences from the data.

The second claim is that the use of theoretical notions indicated above captures a particular kind of abductive inference. This is a reply to van Fraassen [1989], who argues that abductive inference is probabilistically incoherent. To some extent I go along with van Fraassen's way of framing Bayesian abduction. I consider a set of statistical hypotheses, and take their empirical content to be given by their likelihoods. But van Fraassen then proposes to capture the role of theoretical notions, e.g., their explanatory force, by additional changes to the probability assignment over the hypotheses, after processing the Bayesian update. The use of theoretical notions thus leads to probabilistic incoherence. In response, the Bayesian model of abduction proposed in this paper uses theoretical notions to motivate priors over multiple models, which are then adapted by Bayesian conditioning. Because of the different priors, the impact of the data on the models is different, allowing us to tell apart models on the basis of theoretical considerations.

## 2 Bayesian models of abduction

This section discusses some earlier attempts to accommodate the use of theoretical notions in Bayesian inference, and thus to reconcile it with abduction. See, for instance, Day and Kincaid [1994], Okasha [2000], Salmon

[2001], Sober [2002], McGrew [2003], and Lipton [2004].[3] Following the structure of Bayes' rule, these attempts incorporate the explanatory or theoretical considerations in the prior probability, in the likelihoods, or in both.

### 2.1 Abduction by priors

One idea is to model abduction in a Bayesian framework by means of prior probabilities, namely by shifting prior weight to more explanatory hypotheses. However, as argued by Milne [2003], if we want to capture the explanatory considerations in a Bayesian update, then we need to portray this head start of explanatory hypotheses as resulting from the impact of some sort of evidence. Milne then notes that the characteristics that make a hypothesis explanatory, like simplicity, aesthetic quality, and the like, will typically be carried by the hypothesis from the very beginning, because they are logically entailed by the hypothesis or because the hypothesis is constructed to have those characteristics. Therefore, any evidential impact of theoretical characteristics runs into the problem of logical omniscience, or in this case equivalently, the old evidence problem.[4]

There may be theoretical characteristics that are not analytic in this way, but that somehow rely on evidence or background knowledge. But as illustrated by the burglar story in Weisberg [2009], it is still not clear that their impact can then be modeled by means of a prior probability. Say that you find your house in a mess, valuables are missing, and by way of explanation you imagine either of two things: there has been a burglar in the house, or alternatively, one burglar in your house was disturbed by another, then both were discovered by a policeman who chased them away and then took advantage of the situation. Weisberg argues that the first story is more explanatory, quite independently of how probable you find these stories to start with. At least in a subjective Bayesian framework, nothing forces us to align our prior probabilities to our judgement of the explanatory force of the hypothesis.

[3] Douven [1999], Tregear [2004] and Weisberg [2009] react to the criticisms of van Fraassen by showing that explanatory considerations can be modelled as rational changes of belief that do not comply to the Bayesian model. I think these reactions deserve separate attention, but for the moment I seek to maintain the Bayesian norms.

[4] Several authors have proposed Bayesian models of learning such analytic truths; see Earman [1992]. But there is certainly no consensus over these models. I will leave further discussion over them aside.

But even independently of the reasons adduced by Milne and Weisberg, I submit that a Bayesian model of abduction in which explanatory force is defined relative to the data is preferable. That is, a model that involves the likelihoods of the hypotheses is preferred. This allows for the possibility, which seems rather natural, that for certain data one statistical model is more explanatory, while for other data another model is. This is not to say that assigning a high prior to simple or beautiful hypotheses may not be part of a full Bayesian model of abduction. But it seems reasonable to say that such a model cannot be the whole story.

## 2.2 Explanatory likelihoods

A number of authors hold that the likelihoods of hypotheses can be determined, at least partially, by explanatory considerations. Okasha [2000] notes that a high probability of the data given some hypothesis may be motivated by the fact that the hypothesis provides a good explanation of the data. McGrew [2003] presents an interesting elaboration of this idea. He argues that, when updating a hypothesis that entails a particular probabilistic dependence by means of data that confirm this dependence, we effectively exploit the explanatory virtue of consilience.[5]

While I think these are valid responses, I think they will not convince a critic of Bayesian abduction like van Fraassen [1989]. For such a critic, the natural retort is that differences in the likelihood of hypotheses correspond to differences in their empirical content. The argument by van Fraassen against abduction concerns the use of theoretical notions over and above this empirical content. In other words, it concerns the use of abduction for telling apart observationally identical hypotheses. As Weisberg [2009] argues, the employment of high likelihoods for hypotheses with explanatory virtues effectively conflates these virtues with ones that can be expressed probabilistically. The model does not capture what is specifically more ex-

[5] McGrew employs the relevance quotient rather than the likelihoods of the hypotheses because the latter may be hard to determine independently. The salient point for this paper is that the explanatory considerations are tied up with the handling of evidence, rather than being fixed before the evidence comes in. Hence these considerations are expressed in the likelihoods of the hypotheses at issue. In his example concerning the lens hypothesis $L$, we can rephrase the dependence or consilience of the two data $S_1$ and $S_2$ in terms of the likelihoods as $P(S_2|L \cap S_1) = P(S_1|L \cap S_2) \approx 1$, thereby turning the consilience into a fact about likelihoods.

planatory, as opposed to more empirically adequate, about the hypothesis that has in this way gained more posterior probability. In other words, we are looking for a Bayesian model of abduction that somehow manages to do justice to this non-empirical aspect of abduction.

The model of abduction proposed in this paper captures differences in likelihood that are based on explanatory differences in a specific way, namely by means of prior probability assignments over otherwise identical statistical models. I thereby hope to strike a balance between the opposite sides of the debate on Bayesian abduction. I express the explanatory considerations in prior probability assignments over statistical models, but these considerations come out in how the priors interact with the data, namely in the so-called marginal likelihoods. The distinction between the statistical models is theoretical, in the sense that they consist of the same hypotheses. But the theoretical notions motivate different priors, so that they can be distinguished empirically. In sum, both between priors and likelihoods and between theoretical and empirical characteristics, the present model of abduction occupies a middling position.

# 3 Bayesian inference

This section presents the standard Bayesian inference of predictions and parameter values, as discussed in Howson and Urbach [1989], Barnett [1999], Press [2003] and many others. Readers who are familiar with Bayesian statistical inference can skip to the next section.

## 3.1 Tossing coins

Rather unoriginally, I start with an example on coin tossing. This serves as time-honoured stand-in for a much wider set of chance processes concerning independent and identical trials.

Single observations are results of coin tosses, either heads or tails. These are stored as data elements $Q$, tagged with a time index $t$ and a value $q \in \{0, 1\}$ for tails and heads respectively. The result of a single coin toss is thus denoted with $Q_t^q$. For example, if the third coin toss results in heads, we include $Q_3^1$ in the sample. For convenience, sequences of such observations with a length $t$ are denoted with $E^{q_1 \cdots q_t}$, or $E_t$ if the results $q_1 \cdots q_t$ are free variables or clear from the context. For example, if the first tree coin tosses all resulted in heads, we may summarise the data in $E^{111}$ or $E_3$ for short. We can interpret these data elements directly as sets in a possible worlds semantics. The data element $Q^1_3$ is the set of all possible worlds in which the third coin toss indeed results in heads, and similarly $E^{111}$ is the set of all possible worlds in which the first three coin tosses all result in heads. Consequently, we can write $E^{111} = Q^1_1 \cap Q^1_2 \cap Q^1_3$.

Bayesian statistical inference involves statistical hypotheses. In the example on coins, since the tosses are identical and independent, an appropriate statistical hypothesis concerns a constant and independent chance $\theta$ on heads. This chance $\theta$ for the coin to land heads may have a value $1/2$, or $1/3$, or any other real value in the unit interval. The hypotheses $H_\theta$ can be collected in the model $\mathcal{H} = \{H_\theta : \theta \in [0, 1]\}$. Every hypothesis in the model determines a probability assignment over data sets, called the likelihood function of the hypothesis:

$$P(Q^1_{t+1} | H_\theta \cap E_t) = \theta.$$

Given the hypothesis $H_{1/3}$, the chance of observing a toss resulting in heads is $1/3$, the chance of three consecutive heads is $1/27$, and so on.[6] Note that the trials are indeed independent and identical: the likelihoods of the hypotheses $H_\theta$ are such that earlier observations $E_t$ do not matter, and that the likelihoods are independent of the index $t$.

Next to a model, Bayesian statistical inference presupposes a probability function over a statistical model, the so-called prior probability. Note that this is in a sense a second-order probability: it is a probability function whose domain is a set of statistical hypotheses, but these hypotheses are each associated with a probability assignment over the data. In case the model contains a continuum of hypotheses, as in the example, the prior probability can be expressed in a density function, $P(H_\theta)d\theta$. Bayes' theorem may then be used to determine the probability over the model in the light of the observations $E_t$. We obtain a so-called posterior probability as a function of the prior probability and the likelihoods

$$P(H_\theta | E_t)d\theta = \frac{P(E_t | H_\theta)}{P(E_t)} P(H_\theta)d\theta,$$

[6] The statistical hypotheses are not always characterised in terms of conditional probabilities $P(\cdot | H_\theta)$. Classical statisticians prefer the notation $p_\theta(Q^1_{t+1} | E_t) = \theta$.

where the probability of the data set $P(E_t)$ is an average over the likelihoods weighted according to the prior over the hypotheses,

$$P(E_t) = \int_0^1 P(H_\theta)P(E_t|H_\theta)d\theta.$$

Bayesian statistical inference is thus determined by the choice of a model, specifically the likelihood functions of the hypotheses in the model, and by a prior probability over the model. Only the prior and the likelihoods determine the posterior probability over the hypotheses.

### 3.2 Predictions

The posterior probability assignment over the model is the central result of a Bayesian inference. Predictions follow directly from this posterior probability by the law of total probability:

$$\begin{aligned} P(Q^1_{t+1}|E_t) &= \int_0^1 P(Q^1_{t+1}|H_\theta \cap E_t)\, P(H_\theta|E_t)\, d\theta \\ &= \int_0^1 \theta\, P(H_\theta|E_t)\, d\theta. \end{aligned}$$

Note that the latter expression is the expectation value for $\theta$ after observing $E_t$. The predictions based on the hypotheses $H_\theta$ can therefore also be read as Bayesian estimations of the parameter $\theta$.

De Finetti's representation theorem (cf. De Finetti [1937], Paris [1995]) states that the above scheme of hypotheses covers exactly those prediction rules, or estimation functions, that are invariant under order permutations of the observations in $E_t$. We define $t_q$ as the number of $Q^q_i$ in $E_t$. For example, if we observe $q_1 \cdots q_6 = 100111$, then $t_0 = 2$ and $t_1 = 4$, so $t = t_0 + t_1$. The prediction rules resulting from the setup of the example may then be characterised by

$$P(Q^q_{t+1}|E_t) = pr(t_0, t_1).$$

To derive predictions from $E_t$ we only need to know $t_0$ and $t_1$. The order in which the 0's and 1's appear is irrelevant. According to De Finetti's theorem, every rule $pr$ corresponds to a unique prior $P(H_\theta)d\theta$ over the model, and vice versa. For instance, following Festa [1993], if we assume the prior to be a symmetric Beta distribution,

$$P(\theta) \sim \theta^{\lambda/2-1}(1-\theta)^{\lambda/2-1}, \tag{1}$$

we can derive the so-called Carnapian $\lambda$ rules,

$$P(Q^q_{t+1}|E_t) = \frac{t_q + \gamma\lambda}{t + \lambda} = pr_\lambda(t_q, t), \tag{2}$$

in which $\gamma = 1/2$ because the observations are binary (cf. Carnap [1952]). A higher central peak in the density $P(\theta)$ is encoded in a larger parameter $\lambda$. So the larger the value of $\lambda$, the more confident we are that the chance $\theta$ has a value around $1/2$.

In view of this paper's focus, note that the hypotheses $H_\theta$ are in some sense already theoretical. They concern the fixed and independent chance of an observation, and such chances cannot be translated into finite observational terms. Witnessing De Finetti's result, the hypotheses can be eliminated from the inference completely, with the inferences running from data to predictions directly. According to Hintikka [1970], any real empiricist should in fact strive for such an elimination. Nevertheless, as argued in XXX [2003, 2005], there are good reasons for including an intermediate step of statistical hypotheses. The first reason is intelligibility. The hypotheses express the chance mechanism that is assumed to generate the observations, and employing these hypotheses gives us an easy way of turning this assumption into inductive predictions. Moreover, as will be explained below, the hypotheses enable us to express further knowledge of the chance mechanisms in a prior probability over them. It is not always straightforward to incorporate such knowledge in a direct prediction rule.

## 4 Hierarchical Bayesian modeling

The above concerns the use of a single model $\mathcal{H}$. I now direct attention to using two such models together, which differ in virtue of a theoretical notion. In the statistics literature, such a comparison of models falls under the header of Bayesian model selection, or hierarchical Bayesian modeling. For more technical detail, I refer to Gelman and Hill [2007]. Gustafson [2005] provides a detailed discussion of the use of unidentified models, which is particularly salient here. A more philosophical discussion can be found in XXX [2008].

### 4.1 Magical coins

Say that we are going to toss a coin. We are sure that the chances on each trial are independent and identical, so the afore-mentioned hypotheses $H_\theta$ seem appropriate. But imagine that we have further knowledge of the process underlying the data: the coin is either normal, e.g. from an ordinary wallet, or magical, e.g. from a conjurer's box. If the coin is normal, it is most probably fair, having a chance to land heads that is close to $1/2$. And if the coin is magical, it is most probably biased, having a chance to land heads that is close to 0 or 1. On the other hand, the coin may be from my wallet and yet have a highly unusual division of weight, corresponding with a chance away from half. And it may also be from a rather cheap conjurer's box and fail to show the expected bias.

To incorporate this additional knowledge about the coin, we may decide to employ the model concerning constant and independent chances $\theta$ twice. One model may be reserved for the normal coin, $\mathcal{H} = \{H_\theta : \theta \in [0,1]\}$, and another for the magical coin, $\mathcal{H}^\star = \{H^\star_\theta : \theta \in [0,1]\}$. We can distinguish hypotheses $H_\theta$ with different values of $\theta$ within the model $\mathcal{H}$, and similarly we can distinguish $H^\star_\theta$ with different values of $\theta$ within $\mathcal{H}^\star$. But we cannot distinguish the hypothesis $H_\theta$ from the hypothesis $H^\star_\theta$ with the corresponding value of $\theta$, because these hypotheses have identical likelihood functions. Following the above discussion, the distinction between the models $\mathcal{H}$ and $\mathcal{H}^\star$ is theoretical. The combined model $\{\mathcal{H}, \mathcal{H}^\star\}$ is therefore non-identifiable, but in virtue of that we have separate control over the priors defined on the models. Our knowledge concerning the two types of coins motivates specific forms for these priors.[7]

For the sake of simplicity, we choose both functions from the class of symmetric Beta priors, as expressed in Equation (1). For $\lambda = 2$ this distribution is uniform, while larger values $\lambda > 2$ lead to an ever sharper single peak at $\theta = 1/2$, and smaller values $\lambda < 2$ lead to two peaks at $\theta = 0$ and $\theta = 1$, with an ever deeper valley in between. None of these priors is dogmatic, meaning that each of them is nonzero over the whole domain of

[7] If we had no further theoretical story, we would apply something like the principle of indifference or entropy maximization to arrive at a uniform prior over the hypotheses $H_\theta$ in the single model $\mathcal{H}$, or use the non-informative prior devised by Jeffreys [1939]. But in this case, the theoretical background stories motivate different priors for the two submodels.

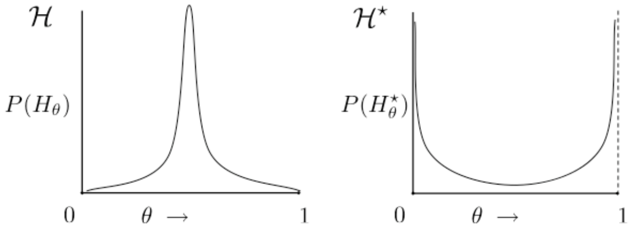

Figure 1: Two different priors over the two models of fixed chance hypotheses, $H_\theta$ and $H^\star_\theta$. The single peak prior is associated with the normal coin, the twin peaks prior with the magical coin.

$\theta \in [0, 1]$. Concretely, to express our expectations for the normal coin, the prior over the model $\mathcal{H}$ may be a symmetric Beta distribution with $\lambda = 12$, so that $P(H_\theta) \sim \theta^5(1-\theta)^5$. This means that we are quite confident that the chance of the normal coin landing heads is close to $1/2$. The prior over the model $\mathcal{H}^\star$ may be $P(H^\star_\theta) \sim (\theta)^{-5/6}(1-\theta)^{-5/6}$, with $\lambda = 1/3$, meaning that we expect the magical coin to have a chance close to 0, or close to 1. These priors are illustrated in Figure 1. Finally, we put a higher-order probability over the models themselves. Since we are initially undecided between the two models, we choose $P(\mathcal{H}) = P(\mathcal{H}^\star) = 1/2$. These assignments together pin down a complete prior over the non-identifiable model.

### 4.2 Bayesian inference over models

We can now apply the Bayesian inference of Section 3. Within the two models, we adapt the probability functions over the statistical hypotheses $H_\theta$ and $H^\star_\theta$ in the light of new data. Each prior separately leads to predictions over coin tosses. The two Beta priors over $\mathcal{H}$ and $\mathcal{H}^\star$ lead to two different Carnapian prediction rules, with $\lambda = 12$ for the normal coin and $\lambda^\star = 1/3$ for the magical one. We write

$$P(Q^q_{t+1}|\mathcal{H} \cap E_t) \quad = \quad pr_{12}(t_q, t), \tag{3}$$

$$P(Q^q_{t+1}|\mathcal{H}^\star \cap E_t) \quad = \quad pr_{1/3}(t_q, t). \tag{4}$$

So the probability assignments within the two models $\mathcal{H}$ and $\mathcal{H}^\star$ may be updated separately, exactly as described in the foregoing.

The probability assignment over the models themselves is also affected by the data set $E_t$. This is the core idea of Bayesian model selection: we may treat the models as separate hypotheses, and run a Bayesian inference on the level of models. This allows us to compute the ratio of posterior model probabilities:

$$\frac{P(\mathcal{H}|E_t)}{P(\mathcal{H}^\star|E_t)} = \frac{P(\mathcal{H})}{P(\mathcal{H}^\star)}\,\frac{P(E_t|\mathcal{H})}{P(E_t|\mathcal{H}^\star)}.$$

In the example the ratio of priors is 1 and hence can be eliminated from the equation. The marginal likelihoods $P(E_t|\mathcal{H})$ and $P(E_t|\mathcal{H}^\star)$ are built up sequentially, using the Carnapian predictions (3) and (4) as likelihoods at each step.

For our example these likelihoods can be given analytical expressions. As indicated, the marginal likelihood is the product of the Carnapian predictions at each step:

$$P(E_t|\mathcal{H}) = \prod_{i=0}^{t_0-1} pr_\lambda(i,i) \prod_{i=t_0}^{t-1} pr_\lambda(i-t_0,i).$$

Notice that the order of the observations does not matter to the marginal likelihoods so that we can reorder them as suggested above. This leads to the expression:

$$\begin{aligned}
P(E_t|\mathcal{H}) \;&=\; \frac{1}{2}\times\frac{\lambda/2+1}{\lambda+1}\times\cdots\times\frac{\lambda/2+t_0}{\lambda+t_0} && (5)\\
&\quad\;\times\frac{\lambda/2+1}{\lambda+t_0+1}\times\cdots\times\frac{\lambda/2+t_1}{\lambda+t-1} && (6)\\
&=\; \frac{(\lambda-1)\text{¡}}{(\lambda+t-1)\text{¡}}\times\frac{(\lambda/2+t_0)\text{¡}(\lambda/2+t_1)\text{¡}}{(\lambda/2-1)\text{¡}\,(\lambda/2-1)\text{¡}} && (7)
\end{aligned}$$

Here $f\text{¡}$ denotes the normal factorial for $f \in \mathbb{N}$ while for $f \in \mathbb{R}\setminus\mathbb{N}$ we have $f\text{¡} = \prod_{i=0}^{\lfloor f\rfloor}(f-i)$, with $\lfloor f\rfloor$ the value of $f$ rounded off. We can derive the same expression using $\lambda^\star$ for the marginal likelihood of $\mathcal{H}^\star$. The ratio of these two expressions gives us the ratio of the marginal likelihoods, or in short the Bayes factor, of the two submodels.

From the posteriors over the non-identifiable model we can subsequently calculate the predictions $P(Q^q_{t+1}|E_t)$. To this aim we weigh the two Carnapian rules with the probabilities of the models. The result is, what Skyrms [1993] calls, a hyper-Carnapian prediction rule:

$$P(Q^q_{t+1}|E_t) = P(\mathcal{H}|E_t)\,pr_{12}(t_q,t) \;+\; P(\mathcal{H}^\star|E_t)\,pr_{1/3}(t_q,t). \qquad (8)$$

The overall prediction may again be taken as an overall Bayesian estimator of the chance for the coin to land heads. It is a mixture of the estimations within the two models.

This setup has some interesting consequences. While the models $\mathcal{H}$ and $\mathcal{H}^{\star}$ consist of pairwise identical hypotheses, the differing priors over them cause different marginal likelihoods. So if we update with a data set for which the relative frequency is close or equal to 0, say $E^{000000}$, we will find that the updated probability of $\mathcal{H}^{\star}$ is larger than that of $\mathcal{H}$. To illustrate this, the table shows a comparison between the effect of $E^{000000}$ on the predictions $P(Q^0_{t+1}|E_t)$, as based on the non-identifiable model $\{\mathcal{H}, \mathcal{H}^{\star}\}$, and as based on the single model $\mathcal{H}$. Over the latter we choose a uniform prior, which leads to the prediction rule $P(Q^0_{t+1}|E_t) = pr_2(t_0, t)$. Also included in the table is the probability assigned to the model $\mathcal{H}^{\star}$ in the non-identifiable case, $P(\mathcal{H}^{\star}|E_t)$.

| Number of observations $t$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Observations $q_t$ | - | 0 | 0 | 0 | 0 | 0 | 0 |
| $P(Q^0_{t+1}\|E_t)$ in $\mathcal{H}$ | 0.50 | 0.67 | 0.75 | 0.80 | 0.83 | 0.86 | 0.88 |
| $P(Q^0_{t+1}\|E_t)$ in $\{\mathcal{H}, \mathcal{H}^{\star}\}$ | 0.50 | 0.71 | 0.79 | 0.85 | 0.90 | 0.93 | 0.94 |
| $P(\mathcal{H}^{\star}\|E_t)$ in $\{\mathcal{H}, \mathcal{H}^{\star}\}$ | 0.50 | 0.50 | 0.62 | 0.73 | 0.81 | 0.87 | 0.91 |

The salient point for us is that the distinction of the two models is theoretical, but that the theoretical notions associated with these models motivate different priors. This making the models empirically distinghuishable.

### 4.3 Discussion

I want to make two further observations about this table, linked to the two central claims of this paper. First, the predictions based on the non-identifiable model are quicker to pick up on the correct relative frequencies. Because the prior over the non-identifiable partition is tailor-made to fit the most likely courses of events it will, if the coin is indeed from a wallet or a conjurer's box, converge to the correct predictions and the true parameter value more quickly. The use of a theoretical notion, i.e. the origin of the coin, thus enables us to improve our predictions. And second, in a direct comparison the model associated with the magical coin becomes more probable than the model associated with the normal coin. This means that we are able to decide over the origin of the coin, a theoretical notion, on the basis of data.

It may seem that the representation of inductive predictions in terms of statistical hypotheses is rather contrived, and that it is much more natural to employ the two models $\mathcal{H}$ and $\mathcal{H}^{\star}$ as hypotheses directly, with the Carnapian rules as their likelihoods. We can then omit the whole story on the underlying models, concentrating on values for $\lambda$ as determinants of the likelihoods instead, and leaving no reason to invoke theoretical notions. But I think there are compelling reasons for spelling out the above in terms of statistical hypotheses. For one, it is hard to make sense of Carnapian prediction rules as the likelihoods of statistical hypotheses. Because such Carnapian hypotheses follow the observed relative frequency, in the limit of sample size $t$ to infinity such hypotheses all have the same likelihood. Hence none of the usual convergence theorems for Bayesian updating applies. The Carnapian hypotheses cannot be said to concern relative frequencies.

I want to add two pragmatic benefits of using the non-identifiable model, relating to the use of hypotheses discussed in Section 3. First, because the two models facilitate the use of differing priors, they enable us to put to use the available knowledge on the nature of the chance process that generates the observations. Second, the use of two models helps us to arrive at intelligible and analytic results. It is in principle possible to work with an informative prior over a single partition $\mathcal{H}$, choosing a single function with the required shape. But the predictions resulting from the combined priors over the two models cannot be equated with a single Carnapian rule, and it is not easy to find some other prediction rule that captures them. It is much more convenient to choose two separate Beta distributions and update these independently. Summing up, the use of a theoretical notion brings a methodological advantage: it makes the statistical inference more transparent, and its use improves the predictive performance.

## 5 Underdetermination and abduction

After discussing the two claims of this paper in the context of the example, we now assess them in more general terms. Below I characterize a particular kind of abductive inference, so-called evaluative empirical abduction, which is arguably captured in the Bayesian model. The section ends with a discussion of the relation between observational and theoretical notions.

## 5.1 Methodological and realist challenges

The problem of underdetermination is that scientific theory is not determined by observation alone: the observations often do not allow us to make a choice between different hypotheses. This presents us with two different challenges, methodological and realist, associated with the afore-mentioned criticisms of Hempel and van Fraassen.

The methodological challenge is to clarify that science is underdetermined in the light of its aims, e.g., to find the truth, or perhaps only the observational truth. The challenge is not that we resolve underdetermination. Rather it is to make sense of the use of underdetermined theoretical structures in science in view of science's goals. The realist challenge, by contrast, is to resolve the underdetermination of science, and thus to safeguard the idea that science provides full epistemic access to the world. A common way of meeting the realist challenge is by defending inference rules such as abduction, which enable us to choose between underdetermined theory on the basis of additional theoretical criteria, like explanatory force. The methodological challenge has attracted far less attention in the literature.

I think that the preceding section presents a partial answer to both of these challenges. Regarding the methodological challenge, the example shows that non-identified statistical models facilitate the expression of suppositions on underlying mechanisms in priors, and that if the suppositions are correct, their use will improve the predictions. That is to say, the statistical inferences are underdetermined, but this underdetermination has a clear methodological use, because it provides us with an opportunity to employ additional knowledge as input to the statistical inference, thereby improving predictive performance. Regarding the realist challenge, the example shows that we can sometimes decide between models that consist of pairwise identical hypotheses. More specifically, theoretical notions may motivate different priors over otherwise identical models, so that these models become observationally distinct. The posterior probability assignment over models is subsequently taken to reflect back onto these theoretical notions. We can thus draw conclusions on typically theoretical notions concerning underlying processes or mechanisms, conclusions that are usually considered to involve abductive inference.

## 5.2 Evaluative empirical abduction

Following the above answer to the realist challenge, we might argue that the foregoing presents us with a Bayesian model of abduction: the examples present a mode of inference allowing us to decide over, in some sense, theoretical notions. The obvious question is: can we call it abductive inference?

Part of the answer is negative. Nothing in the above concerns the generation of suppositions on underlying structure, while many discussions of abduction take that as one of its most important aspects. Moreover, nothing in the above answers to the problem that there are infinitely many theoretical notions that are potentially useful, and that we do not have any reason to choose any particular one when learning from the observations. This problem, which is related to the so-called argument from the bad lot in Van Fraassen [1989], has not been addressed. What the above examples show is that if we happen to choose the theoretical concepts well, then they will be beneficial to the predictions, instead of being irrelevant or detrimental. In other words, the examples show that theoretical notions have an evaluative rather than a generative use. In the model, abduction is thus restricted to the evaluation of a given set of theoretical notions, or to *evaluative abduction* for short.

There is another way in which the above set-up falls short of a proper account of abduction. Concerning the methodological challenge, note that the use of theoretical notions in the examples is directed towards improving the predictions. The above examples have little to say on the use of theoretical notions for improving understanding and explanation, which are equally valid scientific aims. Similarly, concerning the realist challenge, note that we choose between the models not on the basis of theoretical notions proper, but on the basis of the fact that one of the priors, as motivated by the theoretical notions, matches the observations better. So in the end the theoretical notions do have empirical content. Both these aspects are characteristic of the kind of abduction that I take to be captured by the above Bayesian inferences, which may be called *empirical abduction.* We can only show the use of theoretical notions, and model the abductive inferences concerning these theoretical notions, insofar as they lead to certain higher-order empirical expectations, and thus have some observational implications.

### 5.3 Theory-laden observations

The statistical models $\mathcal{H}$ and $\mathcal{H}^\star$ have different empirical content, so it is doubtful that the foregoing captures a mode of inference that tells apart the statistical models on the basis of *theoretical* differences. But notice that theoretical is not the same as non-empirical. Indeed, we can view the above mode of inference as driven by theoretical yet empirical considerations.

We might say that the nature of the distinction between models, as being theoretical or empirical, depends on the content that we take the observation to have. If we emphasize that the models consist of pairwise identical hypotheses and that the contents of the observations is determined completely by the likelihoods of these hypotheses, then the difference between the models is best understood as theoretical. But if we say that the content of an observation is determined by its entire epistemic impact, then the difference between the models is straightforwardly empirical.

In the end the observations allow us to tell the two models apart, and so it is more appropriate to say that the content of the observations is determined by the total effect that updating with the observation has on the probability assignment. But I want to resist the conclusion that, consequently, the above model is completely empiricist in spirit, has nothing to do with theoretical structure, and had better not be called abduction at all. Instead I propose that the content of the observations is partly determined by the prior probability assignments over the hypotheses. That is, the content of observations includes some theoretical content via the theoretical scheme in which we choose to frame these observations, and the observations allow us to draw conclusions on theoretical notions exactly because the observations are framed in this way. Put rather speculatively, the use of non-identified models provides us with a formal account of the theory-ladenness of observations.[8]

In closing, let me relate this to the likelihood principle, according to which the evidential impact of observations is entirely determined by the likelihoods. In parallel to the above considerations, we may ask at what level the likelihood principle is supposed to apply. If it is applied on the level of hypotheses, the models are observationally identical. Applied on the level of models, the models are observationally distinct, but then the

[8] For more on theory-ladenness as a response to the problem of underdetermination, see Okasha [2002].

observations inherit some of the theoretical content from the priors over the models. Whether we like this or not, it seems that in Bayesian model selection the principle must be applied to models, because the theoretically motivated prior over the model is part and parcel of the evidential impact that an observation has.[9]

## 6 Conclusion

I have argued that there is a specific use for underdetermination in statistical inference, and thus in scientific method insofar as it concerns these inferences. I have further argued that in the use of theoretical notions we encounter a particular kind of abductive inference, which I called evaluative empirical abduction. This abductive inference hinges on the interaction between observations and theoretical notions. The statistical inferences of the above provide a formal model of it.

It is tempting to transfer the present insights on evaluative empirical abduction to scientific methodology more generally. The formal model may present an explanation and a justification of the fact that in the face of underdetermination, scientists nevertheless feel that they sometimes have reasons to prefer one theoretical hypothesis over another. They often posit complicated theoretical notions behind relatively poor observational structures. Eventually this will have to be decided by historical case studies and far more detailed analysis, but I suggest that scientists do this in order to allow themselves better ways of using all available knowledge, both in framing observations and in testing theories.

Of course, there are many instances of abductive inference that are not adequately captured by what is presented above. Scientists may well be more creative and speculative than can be accommodated by any probabilistic model of reasoning. In the words of Peter Lipton: their loveliest explanations are often the most unlikely ones. Nevertheless, I hope to have shown that particular concepts can after all be abducted by Bayesians.

[9] Forster [2007] presents arguments against the likelihood principle as a principle for fixing the evidential impact of observations, by showing that different ways of parameterizing lead to different maximum likelihood estimations. The problem cases that Forster comes up with are genuine, but I do not think this reflects badly on the likelihood principle. My hunch is that the problems highlight the evidential import of the prior probability.

## Acknowledgements

Many thanks to Igor Douven, Colin Howson, Franz Huber, Valeriano Iranzo, Elliott Sober, and Leah Henderson for comments on the many earlier versions of this paper. This research was supported by an NWO-Veni grant, the Spanish Ministry of Education, and the Fellowship programme of the Department of Philosophy of the University of Johannesburg.

## References

Barnett, V. [1999] : *Comparative Statistical Inference*, New York: John Wiley.

Carnap, R. [1952] : *The Continuum of Inductive Methods*, Chicago: University of Chicago Press.

Day, T. and H. Kincaid [1994] : 'Putting Inference to the Best Explanation in its Place' in *Synthese* 98 (2), pp. 271–295.

De Finetti, B. [1937] : 'Foresight: its logical laws, its subjective sources' in *Studies in Subjective Probability*, eds. Kyburg, H. and Smokler, H. [1964], New York: John Wiley, pp. 97–158.

Douven, I. [2008] : 'Chapter 27: Underdetermination' in *The Routledge Companion to Philosophy of Science*, eds. Psillos, S. and Curd, M., London: Routledge.

Earman, J. [1992] : *Bayes or Bust*, Cambridge MA: MIT Press.

Earman, J. [1993] : 'Underdetermination, Realism and Reason' in *Midwest Studies in Philosophy XVIII*, eds. P. French et al., pp. 19-38, Notre Dame UP.

Hempel, C. [1958] : 'The theoretician's dilemma' in *Minnesota studies in the philosophy of science II*, Feigl, H., M. Scriven, G. Maxwell (eds.), Minneapolis: University of Minnesota Press.

Howson, C. and Urbach, P. [1989] : *Scientific Reasoning, The Bayesian Approach*, La Salle: Open Court.

Festa, R. [1993] : *Optimum Inductive Methods*, Dordrecht: Kluwer.

Gelman, A. , J. B. Carlin, H. S. Stern, D. B. Rubin [2004] : *Bayesian Data Analysis, Second Edition.* Boca Raton: Chapman and Hall.

Gelman, A. and J. Hill [2007] : Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge: Cambridge UP.

Gigerenzer, G. and D. Goldstein [1996] : 'Reasoning the Fast and Frugal Way: Models of Bounded Rationality', *Psychological Review* 103(4), pp. 650–669

Good, I. J. [1955] : *The Estimation of Probabilities: an Essay on Modern Bayesian Methods*, Cambridge (MA): MIT press.

Gustafson, P. [2005] : 'On Model Expansion, Model Contraction, Identifiability and Prior Information' in *Statistical Science* 20, pp. 111–140.

Hintikka, J. [1970] : 'Unknown Probabilities, Bayesianism, and De Finetti's Representation Theorem' in *Boston Studies in the Philosophy of Science*, Vol. VIII, eds. Buck, R. C. and Cohen, R. S., Dordrecht: Reidel.

Jaynes, E. [1999] : *Probability Theory: The Logic of Science*, accessible at the website http://bayes.wustl.edu/etj/prob.html.

Jeffreys, H. [1939] : *Theory of Probability*, Oxford: Oxford University Press.

Kuipers, T. A. F. [1986] : 'Some Estimates of the Optimum Inductive Method', *Erkenntnis* 24, pp. 37–46.

Lipton, P. [2004] : *Inference to the Best Explanation, second edition*, London: Routledge.

McGrew, T. [2003] : 'Confirmation, Heuristics, and Explanatory Reasoning', *British Journal for the Philosophy of Science* 54, pp. 553–567.

Milne, P. [2003] : 'Bayesianism v. scientific realism', *Analysis* 63, pp. 281–288.

Okasha, S. [2000] : 'Van Fraassens Critique of Inference to the Best Explanation', *Studies in the History and Philosophy of Science* 31(4), pp. 691–710.

Okasha, S. [2002] : 'Underdetermination, holism, and the theory/data distinction', *The Philosophical Quarterly* 52, pp. 303–319.

Paris, J. [1994] : *The Uncertain Reasoner's Companion*, Cambridge: Cambridge University Press.

Pearl, J. [2000] : *Causality*, Cambridge (MA): MIT press.

Press, J. J. [2003] : *Subjective and Objective Bayesian Statistics. Principles, Models, and Applications*, New York: John Wiley.

Psillos, S. [1999] : *Scientific Realism: How Science Tracks Truth*, London: Routledge.

Romeijn, J. W. [2005] : *Bayesian Inductive Logic*, PhD thesis, University of Groningen.

Romeijn, J. W. [2006] : 'Analogical Predictions for Explicit Similarity', *Erkenntnis* 64, pp. 253–280.

Romeijn, J. W. and R. van de Schoot [2008] : 'A Philosophical Analysis of Inequality Constrained Models', in *Bayesian Evaluation of Informative Hypotheses*, Hoijtink, H., I. Klugkist, P. A. Boelen (eds.), New York: Springer.

Royall, R. [2000] : *Statistical evidence : a likelihood paradigm*, London: Chapman and Hall.

Salmon, W. [2001 ] : 'Explanation and Confirmation: a Bayesian Critique of Inference to the Best Explanation', in *Explanations: Theoretical Approaches and Applications*, Hon, G. and S. S. Rakover, Dordrecht: Kluwer, pp. 59–89.

Skyrms, B. [1993] : 'Analogy by Similarity in Hyper–Carnapian Inductive Logic', in *Philosophical Problems of the Internal and External Worlds*, eds. J. Earman, A.I. Janis, G. Massey, and N. Rescher, Pittsburgh: University of Pittsburgh Press, pp. 273–282.

Sober, E. [2002] : 'Bayesianism: its Scope and Limits' in *Bayes' Theorem*, ed. R. Swinburne, Oxford: Oxford University Press, pp. 21–38

Tregear, M. [2004] : 'Utilising Explanatory Factors in Induction?', *British Journal for the Philosophy of Science* 55, pp. 505–519.

Van Fraassen, B. C. [1989] : *Laws and Symmetry*, Oxford: Clarendon Press.

Weisberg, J. [2009] : 'Locating IBE in the Bayesian Framework', *Synthese* 167(1).

Williamson, J. [2003] : 'Bayesianism and Language Change', *Journal of Logic, Language and Information* 12(1), pp. 53–97.