The Review of Symbolic Logic
Volume 15, Number 1, March 2022

# ON THE TRUTH-CONVERGENCE OF OPEN-MINDED BAYESIANISM

TOM F. STERKENBURG
MCMP, LMU Munich
and
RIANNE DE HEIDE
CWI and Leiden University

**Abstract.** Wenmackers and Romeijn [38] formalize ideas going back to Shimony [33] and Putnam [28] into an *open-minded* Bayesian inductive logic, that can dynamically incorporate statistical hypotheses proposed in the course of the learning process. In this paper, we show that Wenmackers and Romeijn's proposal does not preserve the classical Bayesian consistency guarantee of merger with the true hypothesis. We diagnose the problem, and offer a *forward-looking* open-minded Bayesians that does preserve a version of this guarantee.

**§1. Introduction.** On the standard philosophical conception of Bayesian learning, an agent starts out with a particular prior distribution and learns by conditionalizing on the data it receives. Well-known results on the merger of opinion show that the specific prior does not matter too much, as long as there is agreement on what is possible at all. These same results can also be taken to show that the agent converges to the truth, as long as its prior does not exclude this truth from the start [9], p. 141ff, [17].

However, a Bayesian agent cannot include in its prior *every* possible truth from the start; not in practice, and not even in theory [2, 8, 28, 36]. A Bayesian agent must commit to restrictive *inductive assumptions* in its initial choice of prior [16, 30]. Standard results about convergence to the truth only apply if these initial assumptions are actually valid in the learning situation at hand. But there is, on the standard conception, no room for the agent to readjust; not even if these assumptions start looking faulty.

In more explicitly statistical terms, a Bayesian agent's prior can be seen to specify a particular *model*, or set of hypotheses. If the model is appropriate, if one of the hypotheses is *true*, there is—at least for a countable model—a guarantee of *consistency* that the agent with probability 1 (almost surely, a.s.) converges on this truth. But if it is not, the agent's beliefs can with positive probability always and forever remain off the mark. On the standard conception, there is, again, no room for the agent to later adapt this model [7]; there is, in particular, no room to expand the model, to incorporate new hypotheses that might be more in accord with the data [12, 13].

Received: April 22, 2019.
2020 *Mathematics Subject Classification*: 62A01.
*Key words and phrases*: inductive logic, Bayesianism, truth-convergence.




doi:10.1017/S1755020321000022

The question of how to open up the standard conception to make room for incorporating new hypotheses is the Bayesian *problem of new theory* [6], p. 556ff, [9], p. 132f, [31]. An early account that engages with this problem is the *tempered personalism* due to Shimony [33]. Central to Shimony's account is an idea he traces back to Putnam ([28]; see [33], p. 89; [32]), and in more veiled form to Jeffreys ([18]; see [33], p. 97ff; [15]). This is the idea that, rather than taking as starting point an hypothesis set that is as wide as possible, Bayesian inference is relative to a limited set of "seriously proposed hypotheses," that is dynamically expanded as new such hypotheses are proposed. In this context Shimony introduced the notion of a *catch-all hypothesis* that is the complement of all seriously proposed hypotheses at any given time.

Recently, Wenmackers and Romeijn Wenmackers and Romeijn [38] have worked out these ideas, in a statistical setting, into what they brand an *open-minded* Bayesianism. In a number of different versions they propose a Bayesian inductive logic that allows for an agent to adopt newly formulated statistical hypotheses during the learning process.

One important question that they leave untouched, however, is whether these formalizations actually preserve the consistency guarantee of truth-convergence. That is, if the *true* hypothesis is among the actually formulated hypotheses, thus is or becomes part of the open-minded Bayesian's hypothesis set, is the agent from that point on still guaranteed to almost surely converge on this truth? That is the question we investigate in this paper.

We proceed as follows. First, in Section 2, we introduce the statistical framework of Bayesian learning that Wenmackers and Romeijn employ, and discuss their different versions of open-minded agents. Then, in Section 3, we investigate the guarantee of convergence to the truth. We focus on the property of *weak merger* with the true hypothesis, and show that all proposed versions of open-minded Bayesians, unlike the standard Bayesian, *fail* to guarantee this property. In Section 4 we diagnose the problem and the structure of the convergence we could possibly attain, in the course of which we introduce the notions of an *hypothesis* and *posterior scheme* and that of a *completed agent measure*. We then set out for a version of open-minded Bayesianism for which we can show, for every hypothesis and posterior scheme, *strong merger* of the completed agent function, from which weak merger of the agent follows. This is our *forward-looking* open-minded Bayesian. The general threat to truth-convergence lies in the possibility of an endless stream of false overfitting hypotheses: our forward-looking proposal meets this threat by neutralizing the role of old evidence. In an initial proto-version this is achieved by a constraint on the *posteriors* assigned to new hypotheses; in the final version this is achieved by combining a constraint on new hypotheses' *priors* (instantiating the idea of the catch-all) with the stipulation that new hypotheses' likelihoods on past data are equal to the *agent's own past probability assignment*.

We should emphasize that Wenmackers and Romeijn in their paper (and we in this paper) are concerned with the question of how to *incorporate* externally proposed new hypotheses: their proposals are attempts to make this aspect part of a Bayesian logic of inductive inference. They are in their paper (and we are here) *not* concerned with *when* new hypotheses should be taken into consideration, let alone with *how* new hypotheses are conceived. To paraphrase Lindley [25] paraphrasing de Finetti: if you have your statistical model, reasoning is mere calculation, but constructing your model actually requires *thinking*. We are here only concerned with the former, but presume, with Wenmackers and Romeijn, that the scope of mere calculation may be

slightly extended, to the procedure of incorporating given new hypotheses into your model.

**§2. The open-minded Bayesians.** In this section, we first set out the presupposed formal framework (Section 2.1), and then discuss the standard Bayesian (Section 2.2), the *vocal* open-minded Bayesian (Section 2.3), the *silent* open-minded Bayesian (Section 2.4) as well as its *retroactive* variant (Section 2.5), and finally the *hybrid* open-minded Bayesian (Section 2.6).

***2.1. Formal framework: outcomes and hypotheses.*** In the statistical set-up employed by Wenmackers and Romeijn,[1] the domain of a Bayesian agent's probability function is the Cartesian product $\Omega \times \Theta$ of an *outcome space* $\Omega$ and a *statistical hypothesis space* $\Theta$.

*2.1.1. The outcome space.* In all of the following, we assume the simple scenario of repeatedly sampling from two possible elementary outcomes, 0 and 1. Formally, the outcome space $\Omega$ is the space $\{0,1\}^\omega$ of all infinite binary sequences $E^\omega$. It is convenient to treat a probability measure over this space as a function $P$ over the *finite* sequences, that satisfies $P(\varnothing) = 1$, where $\varnothing$ is the empty outcome sequence, and $P(E^t) = P(E^t 0) + P(E^t 1)$ for all finite outcome sequences $E^t$, where $E^t E$ denotes outcome sequence $E^t$ of length $t$ followed by elementary outcome $E \in \{0,1\}$. Formally, the set of *cones* $[\![E^t]\!] := \{E^\omega \in \Omega : E^\omega \text{ extends } E^t\}$ for all finite sequences $E^t$ generates a $\sigma$-algebra $\mathfrak{F}$ over $\Omega$ containing all the Borel sets, and an assignment $P$ as above induces a unique measure $\mu$ on $(\Omega, \mathfrak{F})$ with $\mu([\![E^t]\!]) = P(E^t)$ for all finite $E^t$.

*2.1.2. The hypothesis space.* We consider *statistical* hypotheses that are given by likelihood functions over the possible outcomes. That is, we take hypotheses $H$ to be themselves probability measures over the outcome space.

As a basic example, the i.i.d. or *Bernoulli* hypothesis $H_\theta$ with parameter $\theta \in [0,1]$ assigns each length-$t$ data sequence $E^t$ a probability $H_\theta(E^t) = \theta^{t_1} \cdot (1-\theta)^{t-t_1}$ with $t_1$ the number of 1's in $E^t$. This induces one-step conditional probabilities $H_\theta(1 \mid E^t) = \theta$ at each time point $t$, i.e., independent of the past sequence $E^t$. Thus $H_\theta$ formalizes the data-generating process where the same elementary outcome is always produced with the same probability; for instance, the process of repeatedly tossing a coin (heads is 1, tails is 0) with bias $\theta$.

Other hypotheses can express various dependencies of current probabilities on the structure of the past data. At the extreme end are *deterministic* hypotheses, that at each time only allow for one particular next outcome. This corresponds to a measure assigning probability 1 to (each initial segment of) one particular infinite outcome stream $E^\omega$.

We will assume that at any time the Bayesian agent only has a finite number of explicitly formulated hypotheses. These $N$ hypotheses $H_0, \ldots, H_{N-1}$ are collected in a hypothesis set $\Theta_N := \{H_i\}_{i<N}$.

Below we will consider expanding sequences of hypotheses sets, for which the following notation will be useful. Let $N(t)$ denote the number of hypotheses formulated

[1] For a recent alternative proposal for open-minded Bayesianism in a framework that does not explicitly deal with statistical hypotheses, see Raidl [29].

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

before time $t$, so that the hypothesis formulated at time $t$ (if it exists) is $H_{N(t)}$. We often write $t_0 < t_1 < t_2 < \cdots$ for the time points at which new hypotheses are formulated. In that case we abbreviate $N_i := N(t_i) = N(t_0) + i$, so that $H_{N_i}$ is the hypothesis formulated at $t_i$ and $\Theta_{N_i+1} = \{H_i\}_{i \leq N_i}$ is the hypothesis set right after the formulation of $H_{N_i}$. Note, again, that we do not make any assumptions on the origin of the new hypotheses; all we suppose is that the inquiry prompts some (possibly data-dependent!) stream of incoming hypotheses. We will have more to say about this in our analysis in Section 4.

*2.1.3. Full probability functions from marginal over $\Theta_N$.* Given some distribution over $\Theta_N$ for an agent's marginal probability function over the formulated hypotheses. Since hypotheses are likelihood functions, we can define the agent's marginal likelihood function over the outcomes, *conditional* on hypothesis $H_i$, by

$$P(E \mid H_i) := H_i(E). \tag{1}$$

Then by the law of total probability we obtain the unconditional marginal likelihood over the outcomes by

$$P(E) = P(E \mid \Theta_N) = \sum_{i<N} P(H_i) \cdot H_i(E). \tag{2}$$

Thus stipulating the marginal over $\Theta_N$ defines a probability function $P$ over all of $\Omega \times \Theta_N$.[2]

***2.2. The standard Bayesian.*** A Bayesian agent starts with a set $\Theta_N$ of $N$ hypotheses, and a probability function $P_0$, or *prior*, over $\Theta_N$ and hence over $\Omega \times \Theta_N$.[3] When the agent receives a new outcome $E_t$ at time $t > 0$, it must update its probability function $P_{t-1}$ to a new probability function or *posterior* $P_t$.

The orthodox Bayesian way of updating on the evidence is by use of *Bayes's rule*,

$$P_t(\cdot) := P_0(\cdot \mid E^t), \tag{3}$$

with $E^t$ the outcome sequence up to time $t$. In particular, for the agent's *predictive probabilities*, or its marginal probability function over finite-length future outcomes,

$$P_t(E^s) = P_0(E^s \mid E^t) = \frac{P_0(E^t E^s)}{P_0(E^t)}. \tag{4}$$

[2] Our account of hypotheses is a slightly simplified version of Wenmackers and Romeijn's. They take as hypotheses *sets* of probability functions, so that there is a difference between the "theoretical context" $T_N = \{H_i\}_{i<N}$, the set of hypotheses, and $\Theta_N = \cup_{i<N} H_i$, the set of all probability functions that constitute the hypotheses. Furthermore, an hypothesis's likelihood is then only settled with the aid of a subprior over the hypothesis's elements. While this additional complexity arguably does more justice to the actual shape of hypotheses in scientific or statistical inference, nothing in the following should hinge on the simpler formulation we have chosen to adopt. (Wenmackers and Romeijn's running example of the food inspection also only figures "elementary" hypotheses that are singleton sets, i.e., single probability functions as in our framework.) That said, a natural further development of the current work would allow for representing "hypotheses" as models in the form of continuous distributions over parametric hypothesis spaces, so as to be able to explicitly analyze, for instance, adding more parameters to a model.

[3] We always assume that the prior for given hypothesis set $\Theta_N$ is *regular*, meaning that it assigns nonzero probability to each element in $\Theta_N$.

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

Equivalently but more in line with the procedure in Section 2.1.3, the agent first updates the marginal posterior over the hypotheses, again by Bayes's rule and by Bayes's theorem:

$$P_t(H_i) := P_0(H_i \mid E^t) = \frac{P_0(H_i) \cdot H_i(E^t)}{P_0(E^t)}. \tag{5}$$

Then, by the law of total probability on the conditional marginal likelihood,

$$P_t(E^s) = P_0(E^s \mid E^t) = \sum_{i<N} P_t(H_i) \cdot H_i(E^s \mid E^t). \tag{6}$$

Note that $P_t$ is here always implicitly conditional on the hypothesis set $\Theta_N$, e.g., $P_t(H_i) = P_t(H_i \mid \Theta_N)$.

In summary, the **standard Bayesian** proceeds as follows.

$(t = 0)$ *N hypotheses.* At the start each explicitly formulated hypothesis $H_i$ in $\Theta_N$ receives a prior $P_0(H_i) > 0$, such that $\sum_{i<N} P_0(H_i) = 1$.

$(t > 0)$ *Evidence* $E^t$. Updating on evidence at a later point in time proceeds by Bayes's rule, equation (5).

$(t > 0)$ *New hypothesis* $H_N$. An hypothesis formulated at a later point in time is not an element of the set $\Theta_N$ of hypotheses. This hypothesis's prior and posterior probability is and will always remain 0.

***2.3. The vocal open-minded Bayesian.*** Wenmackers and Romeijn's proposal of an open-minded Bayesianism starts with postulating, alongside the set $\Theta_N$ of explicitly formulated hypotheses, a *catch-all hypothesis* ([38], p. 1231f; an idea presented in but preceding [33], p. 95; e.g., Savage in [1], p. 70). This catch-all hypothesis $\overline{\Theta}_N$ comprises all (yet) unformulated hypotheses; Wenmackers and Romeijn explicitly define it as the complement of $\Theta_N$ within the class of all possible hypotheses.

Their *vocal* variant of open-minded Bayesianism ([38], p. 1234f, 1238ff) derives its name from the fact that the catch-all hypothesis comes with a symbolic prior and likelihood function that figures in all calculations. This in contrast to the *silent* version (Section 2.4), where no such prior or likelihood is formulated.

*2.3.1. Specification.* Thus the vocal open-minded Bayesian starts with an hypothesis set $\Theta_N$ of $N$ explicitly formulated hypotheses, and in addition a catch-all hypothesis $\overline{\Theta}_N$. Each explicit hypothesis is assigned a numerical prior probability, summing to 1; and in addition the catch-all hypothesis is assigned an "indefinite" or "merely symbolic" prior $\tau_N$. The numerical probability assigned to an $H \in \Theta_N$ specifies the prior probability value $P_0(H \mid \Theta_N)$, *conditional* on the hypothesis set; the *unconditional* or absolute prior is given by the normalization $P_0(H) := (1 - \tau_N) \cdot P_0(H \mid \Theta_N)$, which is also indefinite because it involves $\tau_N$. While the catch-all thus receives an explicit yet indefinite prior value $P_0(\overline{\Theta}_N) = \tau_N$, the prior probability values $P_0(H')$ of the (yet) unformulated hypotheses $H' \in \overline{\Theta}_N$ are left fully unspecified.

In addition to the indefinite prior, the catch-all comes with a symbolic likelihood function $x_N(\cdot) := P_0(\cdot \mid \overline{\Theta}_N)$. Thus the unconditional marginal likelihood function,

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

analogous to (2) but now not conditional on $\Theta_N$, is given by the indefinite term

$$P_0(E) = \sum_{i<N} P_0(H_i) \cdot H_i(E) + \tau_N \cdot x_N(E)$$
$$= (1 - \tau_N) \sum_{i<N} P_0(H_i \mid \Theta_N) \cdot H_i(E) + \tau_N \cdot x_N(E).$$

The calculation of an explicit hypothesis's posterior on receiving evidence $E$ proceeds by Bayes's rule and theorem in accordance with (5), but now also results in an indefinite term because it involves $P_0(E)$.

Finally and crucially, at each point in time the open-minded Bayesian may receive a *newly formulated hypothesis*. This new hypothesis, in terminology due to Earman ([9], 196), is *shaved off* from the catch-all. Formally, the vocal agent extends its current hypothesis set $\Theta_N$ to the new set $\Theta_{N+1} = \Theta_N \cup \{H_N\}$ to include the newly formulated hypothesis $H_N$, leaving a cleanly shaven catch-all $\overline{\Theta}_{N+1} = \overline{\Theta}_N \setminus \{H_N\}$. To specify the new hypothesis's prior $P_0(H_N)$ the agent then chooses a prior probability value $p$ that it takes from the prior $\tau_N$, leaving the indefinite remainder $\tau_{N+1} := \tau_N - p$ for the new catch-all $\overline{\Theta}_{N+1}$. Writing $x_{N+1}(\cdot) = P_0(\cdot \mid \overline{\Theta}_{N+1})$ for the new catch-all's indefinite likelihood function, expressions for the marginal likelihoods and posteriors that explicitly contain $H_N$ can be calculated as above.

In summary, the **vocal open-minded Bayesian** proceeds as follows.

($t = 0$) *$N$ explicit hypotheses.* Each explicit hypothesis $H_i$ in $\Theta_N$ receives a prior $P_0(H_i \mid \Theta_N) > 0$ conditional on $\Theta_N$, such that $\sum_{i<N} P_0(H_i \mid \Theta_N) = 1$. Moreover, the catch-all hypothesis $\overline{\Theta}_N = \Theta \setminus \Theta_N$ receives an indefinite unconditional prior $P_0(\overline{\Theta}_N) := \tau_N$, and the unconditional priors of the explicit hypothesis are given by $P_0(H_i) := (1 - \tau_N) \cdot P_0(H_i \mid \Theta_N)$.

($t > 0$) *Evidence $E^t$.* Updating proceeds in the standard fashion, although involving an indefinite prior and likelihood of the catch-all:

$$P_t(H_i) := P_0(H_i \mid E^t) = \frac{P_0(H_i) \cdot H_i(E^t)}{\sum_{j=0}^{N-1} P_0(H_j) \cdot H_j(E^t) + \tau_N \cdot x_N(E^t)}.$$

($t > 0$) *New hypothesis $H_N$.* When a new explicit hypothesis $H_N$ is formulated, extending the hypothesis set to $\Theta_{N+1} = \Theta_N \cup \{H_N\}$, the prior $\tau_N$ of the earlier catch-all is decomposed into a value $p < \tau_N$ for the prior $P_0(H_N)$ of the new hypothesis and a remainder $\tau_{N+1} = \tau_N - p$ for the prior $P_0(\overline{\Theta}_{N+1})$ of the new catch-all.

*2.3.2. Discussion.* The obvious drawback of this proposal is the introduction of purely symbolic terms for the priors and likelihoods of the catch-all. Apart from the pain of doing actual calculations with these terms, it is quite unclear how to understand them.

Wenmackers and Romeijn variously refer to these terms as "unknown," "undefined," "indefinite," or "unspecified." But even if we grant that these terms can be considered unknown to the agent (leaving aside worries about the notion, not just of an unknown probability, but of an unknown *epistemic* probability), it seems to us that there is a difference between terms that are unknown yet *definite*, and terms that are not. Only

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

in the first case is there an actual matter to the fact of, say, $\tau_N < c$ for a numerical constant $c$. Thus it is only in the first case that it is clear that the shaving off from the catch-all actually imposes a *limitation* on how much prior the agent can still assign to a newly formulated hypothesis.[4] In contrast, it is less clear whether an *indefinite* probability value allows for shaving off *any desired* definite prior. This might not be a problem to Wenmackers and Romeijn; indeed this would fit their suggestion that the unconditional probability of the catch-all's complement is always *infinitesimally* small (ibid., 1248). However, for our purposes it will prove to be important to impose such constraints on the agent, which is why we will not further pursue the idea of indefinite or infinitesimal priors.

***2.4. The silent open-minded Bayesian.*** The motivation for the *silent* version of open-minded Bayesianism ([38], p. 1234f, 1241f) is to evade the difficulties surrounding symbolic assignments to the catch-all. This is achieved by doing away with this assignment altogether, namely, by always only considering *conditional* evaluations, conditional on the current hypothesis set. The corresponding agent is simply silent about the *absolute* probability values.

*2.4.1. Specification.* The silent open-minded Bayesian starts out, as before, with an hypothesis set $\Theta_N$ of explicitly formulated hypotheses, assigning each $H \in \Theta_N$ a conditional probability value $P_0(H \mid \Theta_N)$. As opposed to the vocal open-minded Bayesian, there is no bookkeeping of the catch-all or the unconditional prior $P_0$.

Since all probability terms are conditional on the current hypothesis set, updating on evidence proceeds fully conditional on $\Theta_N$. That is, the term $P_t(H_i \mid \Theta_N)$ is evaluated via the usual Bayesian updating (5), conditional on $\Theta_N$.

If a new hypothesis $H_N$ is formulated, the silent open-minded Bayesian again extends its current hypothesis set $\Theta_N$ to the new set $\Theta_{N+1} = \Theta_N \cup \{H_N\}$ to include the newly formulated hypothesis $H_N$. It then assigns the new hypothesis conditional on the new hypothesis set a *posterior* value of choice, i.e., a value for $P_t(H_N \mid \Theta_{N+1})$. The new posterior values of the earlier hypotheses are calculated by renormalization, thus preserving the probability ratios.

In summary, the **silent open-minded Bayesian** proceeds as follows.

$(t = 0)$ *$N$ explicit hypotheses.* Each explicit hypothesis in $\Theta_N$ receives a prior $P_0(H_i \mid \Theta_N)$ conditional on the initial hypothesis set.

$(t > 0)$ *Evidence $E^t$.* Updating proceeds in the usual way, conditional on the current hypothesis set $\Theta_N$:

$$P_t(H_i \mid \Theta_N) := P_0(H_i \mid E^t, \Theta_N) = \frac{P_0(H_i \mid \Theta_N) \cdot H_i(E^t)}{P_0(E^t \mid \Theta_N)}.$$

$(t > 0)$ *New hypothesis $H_N$.* When a new hypothesis $H_N$ is formulated, extending the hypothesis set to $\Theta_{N+1} = \Theta_N \cup \{H_N\}$, the posterior $P_t(H_N \mid \Theta_{N+1})$ is set to a value $p \in (0, 1)$, and the posteriors of the remaining explicit hypotheses conditional on the new hypothesis set are renormalized by

$$P_t(H_i \mid \Theta_{N+1}) := (1 - p) \cdot P_t(H_i \mid \Theta_N).$$

[4] For instance, Wenmackers and Romeijn [38], p. 1240 mention the possibility of assigning a uniform prior to a new hypothesis. If $\tau_N$ has an (unknown yet) definite value, then that would only be possible if this value is in fact greater than $\frac{1}{N+1}$.

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

*2.4.2. Discussion.* In the silent version Wenmackers and Romeijn do away with the explicit monitoring of the catch-all hypothesis by simply always "hiding behind the conditionalization stroke" ([38], p. 1243). As they themselves point out, one might feel uneasy about thus leaving unspecified the agent's unconditional, *absolute* convictions. One might indeed feel that this threatens to sufficiently compromise coherence that this is no longer a Bayesian account (cf. [14], p. 1282).

However, it is surely more in tune with statistical practice that probabilities are always evaluated under the tentative assumption of a particular model, without any pledge to the truth of this model. The discussion by Sprenger ([34], also see [35], Chap. 12, [37]) is a recent example of several earlier expressions of this view in the Bayesian literature (e.g., [25], p. 334; [24], p. 611), that tends to go together with a commitment to coherence only for as long as the model does not change (see indeed [33], p. 103f). Perhaps most outspoken in this latter respect is Howson's account of Bayesianism, "a theory of valid inductive inference from pre-test to post-test distributions," which offers the worry of an "inconsistent assignment over time" a simple reply: "so what?" ([15], p. 81).

Moreover, while there is a clear sense in which open-minded Bayesianism, when moving to larger models, loses the guarantee of *dynamic* coherence (see Section 4.1.3 for more details), there is also an important sense in which Wenmackers and Romeijn's proposals are "conservative extensions" where the probabilities conditional on an expanded model do cohere with those conditional on the original model ([38], p. 1235f). Bayes's rule amounts to restricting the subalgebra on the outcome space (to the subtree of the outcome space that extends the evidence) while preserving all probability ratios within; the rule for incorporating new hypotheses *enlarges* the subalgebra on the *hypothesis space* (to the larger hypothesis set) while likewise preserving all probability ratios within the original (ibid.).

We conclude that the silent version holds a conceptual advantage over the vocal version. The main *formal* difference, for our purposes, is that in the vocal version, a new hypothesis is assigned a certain prior value that is constrained by the catch-all's prior; whereas in the silent version, a new hypothesis is assigned a *posterior* value, the choice of which is *unconstrained*.

Wenmackers and Romeijn indeed worry that "[t]he silent proposal allows too much freedom in the assignment of a posterior to the new hypothesis—so much freedom, that it is not clear that the old evidence has any impact" (ibid., 1245). This prompts them to propose a *hybrid* variant of the vocal and the silent versions (Section 2.6). Before we turn to this version, we will take a quick look at a more direct tweak of the silent version that replaces the choice of posterior by the choice of prior, so that the calculation of the former requires some "reconstructive work" that does take old evidence into account (ibid., 1242).

***2.5. The silent open-minded Bayesian: retroactive variant.*** Thus the alternative variant of the silent version is where we "retroactively" assign a *prior* to a new hypothesis, i.e., a value $p_0$ to $P_0(H_N \mid \Theta_{N+1})$. After renormalizing the priors of the other hypotheses,

$$P_0(H_i \mid \Theta_{N+1}) := (1 - p_0) \cdot P_0(H_i \mid \Theta_N) \tag{7}$$

for all $H \in \Theta_N$, we can with the help of Bayes's rule (using the the new likelihood $H_N(E^t)$), calculate $P_t(H_N \mid \Theta_{N+1})$ from there.

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

Note, however, that for the end result it does not make a difference whether we calculate a posterior from a choice of prior, or the other way around. (Provided that $H_N$'s likelihood on $E^t$ is positive, or its posterior can only be 0.) For any posterior $p_t$ for a new hypothesis, we can uniquely reconstruct a choice of prior $p_0$ that in combination with the new hypothesis's likelihood, will result after $E^t$ in *that* posterior. After all, there are, unlike in the vocal version, no constraints on choosing a prior $p_0$.

***2.6. The hybrid open-minded Bayesian.*** The vocal and the silent version are combined in the *hybrid* version ([38], p. 1245f) as follows. The agent starts out, as in the vocal version, with an explicit assignment to the catch-all hypothesis. During the normal learning process of updating on the evidence, it stays in the "silent phase," in which it evaluates all probabilities conditional on the current hypothesis set. Only when a new hypothesis is formulated does it enter the "vocal phase," in which it, like in the vocal version, retroactively shaves off a prior for the new hypothesis from the catch-all's prior, after which it, like in the retroactive silent version, recalculates the priors and posteriors (again conditional, but on the *new* hypothesis set) from there.

In summary, the **hybrid open-minded Bayesian** proceeds as follows.

$(t = 0)$ *$N$ explicit hypotheses.* Each explicit hypothesis $H_i$ in $\Theta_N$ receives a prior $P_0(H_i \mid \Theta_N) > 0$ conditional on $\Theta_N$, such that $\sum_{i<N} P_0(H_i \mid \Theta_N) = 1$. Moreover, as in the vocal version, the catch-all hypothesis $\overline{\Theta}_N = \Theta \setminus \Theta_N$ receives an unconditional prior $P_0(\overline{\Theta}_N) := \tau_N$, and the unconditional priors of the explicit hypothesis are given by $P_0(H_i) := (1 - \tau_N) \cdot P_0(H_i \mid \Theta_N)$.

$(t > 0)$ *Evidence $E^t$.* Updating proceeds as in the silent version, conditional on the current hypothesis set $\Theta_N$:

$$P_t(H_i \mid \Theta_N) := P_0(H_i \mid E^t, \Theta_N) = \frac{P_0(H_i \mid \Theta_N) \cdot H_i(E^t)}{P_0(E^t \mid \Theta_N)}.$$

$(t > 0)$ *New hypothesis $H_N$.* When a new explicit hypothesis $H_N$ is formulated, extending the hypothesis set to $\Theta_{N+1} = \Theta_N \cup \{H_N\}$, as in the vocal version the unconditional prior $\tau_N$ of the earlier catch-all is decomposed into a value $p < \tau_N$ for the unconditional prior $P_0(H_N)$ of the new hypothesis and a remainder $\tau_{N+1} = \tau_N - p$ for the unconditional prior $P_0(\overline{\Theta}_{N+1})$ of the new catch-all. The priors conditional on the new hypothesis set are obtained by renormalization,

$$P_0(H_i \mid \Theta_{N+1}) = \left(1 - \frac{p}{1 - \tau_{N+1}}\right) \cdot P_0(H_i \mid \Theta_N),$$

from which the conditional posteriors are obtained by the usual updating,

$$P_t(H_i \mid \Theta_{N+1}) := P_0(H_i \mid E^t, \Theta_{N+1}) = \frac{P_0(H_i \mid \Theta_{N+1}) \cdot H_i(E^t)}{P_0(E^t \mid \Theta_{N+1})}.$$

Thus the hybrid version combines the conceptually more pleasing conditional reasoning of the silent version with the constraint on new priors introduced by the catch-all in the vocal version. This constraint proves important for our concern in this paper, the guarantee of truth-merging.

**§3. The failure of truth-convergence.** We start by introducing the formal property of convergence to the truth, as satisfied by the standard Bayesian (Section 3.1). After

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

some preliminary remarks about this property in the open-minded case (Section 3.2), we demonstrate and diagnose its failure for the silent (Section 3.3) and the hybrid (Section 3.4) version.

***3.1. The standard Bayesian.*** Suppose the standard, "closed-minded" Bayesian starts with a hypothesis set that includes the hypothesis $H^*$ that is actually *true*, meaning that the probabilities given by $H^*$ are the true probabilities that govern the generation of the data. In that case, one can prove a strong statement about the agent's convergence to this truth. Namely, one can prove that, $H^*$-almost surely, the *total variational distance*

$$\sup_{A\in\mathfrak{F}} |P_t(A) - H^*(A \mid E^t)| \tag{8}$$

between the agent's probabilities and the $H^*$-probabilities on future events goes to 0 as $t \to \infty$. That is, *with true probability* 1 (as given by $H^*$), the agent's probabilities conditional on the past will convergence on all events' *true* probabilities. We say that the agent *strongly merges* with the truth.

DEFINITION 3.1. *For probability measures $P$ and $Q$ on $(\Omega, \mathfrak{F})$, we say that $P$ strongly merges with $Q$ if $Q$-a.s.*

$$\sup_{A\in\mathfrak{F}} |P(A \mid E^t) - Q(A \mid E^t)| \xrightarrow{t\to\infty} 0. \tag{9}$$

A standard Bayesian's strong merger with the truth follows directly from a fundamental result due to Blackwell and Dubins.

THEOREM 3.2 [3]. *For probability measures $P$ and $Q$ on $(\Omega, \mathfrak{F})$ such that the latter is absolutely continuous with respect to the former, i.e., $Q(A) > 0$ implies $P(A) > 0$ for all events $A$ in the $\sigma$-algebra $\mathfrak{F}$ on $\Omega$, it holds that $P$ strongly merges with $Q$.*

Namely, if the Bayesian agent's hypothesis set contains $H^*$, meaning that its regular prior probability $P(H^*) > 0$, then, in terminology due to Kalai and Lehrer [19, 20], *$P$ holds a grain* of $H^*$, or *$P$ holds a grain of the truth*. That is to say, there is an $a \in (0, 1]$, namely $a = P(H^*)$, such that the marginal prior *$P$ on the outcome space* equals $a \cdot H^* + (1 - a) \cdot P'$, for some other measure $P'$. More precisely still, from the fact that $P(H^*) > 0$, we have that *$P$ dominates $H^*$*, meaning that $P(E^t) \geq a \cdot H^*(E^t)$ for all finite outcome sequences $E^t$. But one can show that this implies that also $P(A) \geq a \cdot H^*(A)$ for *all* events $A \in \mathfrak{F}$ generated from the finite sequences, which just means that $H^*$ is absolutely continuous with respect to $P$.

COROLLARY 3.3. *If $P$ holds a grain of the truth $H^*$, then $P$ strongly merges with $H^*$.*

Strong merger is a very strong notion, as it includes all tail events $A$, the occurrence of which can never be verified. A more down-to-earth notion of truth-convergence is *weak merger* [20], that only concerns the special case of the next outcome. This is the notion we will be focusing on in this paper.

DEFINITION 3.4. *For probability measures $P$ and $Q$ on $(\Omega, \mathfrak{F})$, we say that $P$ weakly merges with $Q$ if $Q$-a.s.*

$$\sup_{E_{t+1}\in\{0,1\}} |P_t(E_{t+1}) - H^*(E_{t+1} \mid E^t)| \xrightarrow{t\to\infty} 0. \tag{10}$$

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

In fact, weak merger of two probability measures is equivalent, for every $d \in \mathbb{N}$, to merger where the supremum ranges over all future outcomes of length up to $d$ (ibid.). Nonetheless, as we will explain in more detail in our analysis in Section 4, we will in this paper focus on the case $d = 1$. Moreover, as we will explain there too, the notion of holding a grain of the truth, already sufficient for strong merger, will be central to our analysis. When in the following we refer to "truth-convergence" without further qualification, we mean weak merger as in Definition 3.4.[5]

***3.2. The open-minded Bayesians.*** The question we shall investigate is whether Wenmackers and Romeijn's proposals can retain this conception of convergence to the truth, *whenever the true hypothesis $H^*$ is formulated*. More precisely, the question is whether we can show that, if $H^*$ is indeed formulated at some time $t_0$, the agent function $P_t(\cdot \mid \Theta_{N(t)})$, as $t > t_0$ goes to infinity, weakly merges with $H^*$. The question is whether we can show that, after $H^*$ has been formulated, a.s.

$$\sup_{E_{t+1} \in \{0,1\}} |P_t(E_{t+1} \mid \Theta_{N(t)}) - H^*(E_{t+1} \mid E^t)| \xrightarrow{t \to \infty} 0. \tag{11}$$

One might already object here that we should rather consider convergence of the *unconditional* agent function $P_t(\cdot) = P_t(\cdot \mid \Theta_{N(t)} \cup \overline{\Theta}_{N(t)})$. For an adherent to the vocal variant, the agent's beliefs are constituted by a function over all hypotheses, including those in the catch-all, and so an agent's truth-convergence should be taken to mean convergence of *that* function. However, we already argued in favour of the conditional perspective of the silent or hybrid version; and it is unclear to us where to even start in analyzing convergence of a measure that is partly unspecified.

This is not to say that the truth-merging of $P_t(\cdot \mid \Theta_{N(t)})$ is unproblematic in its interpretation. Indeed, we will below be much concerned with meeting two challenges in squaring the semi-formal expression (11) with our intuitive demand of truth-convergence. *Semi*-formal, because we have not yet made fully clear how to understand the agent functions, with their dependence on proposed hypotheses, under the scope of an "almost-sure" qualifier on outcome streams. For this we will need to get clear, first, on the exact nature of the probability-1 qualification, and second, on the exact nature of the agent measure that we seek merging for.

Nevertheless, the demand that (11) is supposed to capture is already sufficiently precise to isolate a straightforward case in which truth-convergence *is* guaranteed (Section 3.2.1). This will then also already point us to the general case that might be problematic (Section 3.2.2). In fact, this is already enough to show that this case *is*

[5] There exist other notions of truth-convergence one could consider. Note, first of all, that the presupposition of a true *statistical hypothesis* can be distinguished from what is perhaps the more usual setting in philosophy, where the relevant truths are the events or elements of the outcome space [9, 11]. Note, further, that the notion of merging is concerned with learning the probabilities *of future outcomes*. This can be distinguished from learning *the correct hypothesis* ("learning the parameter" in a statistical model), which would correspond to the agent's posterior concentrating on the correct element in the hypothesis set. One reason why we do not consider this notion here is that such posterior-concentration is rather trivially impossible unless we exclude the possibility of different hypotheses that from some point on are "empirically equivalent" in that they give the same predictive probabilities (cf. [22], p. 148f). Finally, there are still less powerful notions of truth-merging, including *almost weak merging*. See [22] and [23] for overviews of learning notions and necessary and sufficient conditions.

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

problematic: all the variants of open-minded Bayesianism are *not* in general guaranteed to preserve truth-convergence (Sections 3.3–3.4). Only in the discussion leading up to our diagnosis of this failure and our proposal of a *forward-looking* open-minded Bayesian, in Section 4, will we finally face the aforementioned challenges head-on.

*3.2.1. Finitely many new hypotheses.* The answer to our question is a clear *yes* if we can be sure that, after $H^*$ is formulated, *no further new hypotheses will ever be formulated*. For each of the different versions of open-minded Bayesianism, the agent with function $P_t(\cdot \mid \Theta_{N(t)})$ after formulation of $H^*$ can then be treated as a standard Bayesian that starts its investigation at $t$ with a fixed hypothesis set $\Theta_{N(t)}$. Thus, as $H^* \in \Theta_{N(t)}$, the agent then holds a grain of the truth and we can simply apply Corollary 3.3 to $P_t(\cdot \mid \Theta_{N(t)})$ to indeed obtain not just weak but strong merger with the truth from there.

This observation easily extends to the more general case where we can be sure that after some finite point in time there will no longer be new hypotheses formulated. So suppose $H^*$ is formulated at $t_0 \leq t$, say in response to data $E^{t_0}$. Then, to put it graphically, from each of the possible nodes $E^t$ in the outcome tree extending $E^{t_0}$, we can run Corollary 3.3 on the fixed agent function to obtain, with probability 1, truth-convergence from *there*; but that means we already have the guarantee of truth-merger from *here*, at $E^{t_0}$. Hence, under the assumption that no more hypotheses are formulated after some finite time $t$, we have strong merger whenever the truth $H^*$ is formulated. This assumption can be reformulated as saying that, on any infinite outcome stream, only a finite number of new hypotheses will ever be formulated.

FACT 1. *All open-minded Bayesians are guaranteed to strongly merge with the truth whenever the truth is formulated, if there is a finite bound on the number of new hypotheses that will be formulated.*

*3.2.2. Infinitely many new hypotheses.* The previous assumption, in entailing that from some point on the open-minded Bayesian reduces to a standard, fixed-minded, Bayesian, thereby also neutralizes a good part of the distinctive interest of the former. It is, more importantly, an assumption that we do not generally want to make: we certainly do not want to assume that, when the true hypothesis is formulated, who or whatever is responsible for designing new hypotheses *knows* that it can stop now.

On the other hand, it also sounds unrealistic that in an actual scientific inquiry, certainly after the true hypothesis has already been found, one would mindlessly keep incorporating newly arriving hypotheses indefinitely. One would presumably only look out for new hypotheses if the currently available ones do not seem to do: if there is some misfit between the data and the current hypotheses. Incorporating this element, possibly in the shape of a formal model verification procedure, would still not render the scenario of an unending stream of false hypotheses insignificant: there is now a tension to be resolved between risking sticking to suboptimal hypotheses and risking incorporating false ones.

Important as this element is, it is beyond the scope of the current paper. We are here first concerned with the consistency requirement of truth-convergence in the most general case where the agent might forever keep receiving new (and false) hypotheses, which it has to incorporate irrespective of the past outcomes and current hypothesis set.

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

This general case is potentially problematic because if the agent keeps having to distribute some of its posterior to these new and false hypotheses (and so keeps having to incorporate these in its predictions), this could get in the way of its converging on the true hypothesis's *true* predictive probabilities. In fact, this *is* problematic, for all the versions of open-minded Bayesianism. We now first look at the silent variants (Section 3.3), where this shows very directly; and then at the more interesting hybrid variant (Section 3.4).

***3.3. The silent open-minded Bayesian.*** This version is the least constrained of the open-minded Bayesianisms, which makes it most straightforwardly fail to guarantee truth-convergence. We first show this for the standard open-minded version of Section 2.4, and then for the retroactive variant of Section 2.5.

*3.3.1. The silent open-minded version: original variant.* The reason for the failure of truth-convergence is that we cannot exclude infinite streams of false hypotheses that keep occupying a specific share of the posterior probability and in this way keep distorting the predictive probabilities.

FACT 2. *The original variant of the silent open-minded Bayesian is not guaranteed to weakly merge with the truth whenever the truth is formulated.*

EXAMPLE 3.5. *Consider the scenario where the data is generated by some Bernoulli measure $H_{\theta^*}$. Suppose for concreteness that $\theta^* = 9/10$, and that this true hypothesis $H^* = H_{\theta^*}$ is indeed formulated at some stage $t_0$. Now consider the possibility that infinitely often (i.e., for each stage $t' > t_0$ there is a still later stage $t > t'$ at which) a new hypothesis $H_{N(t)}$ is formulated that issues a one-step predictive probability $H_{N(t)}(1 \mid E^t) = 0$. Since there are no restrictions on the posterior which the silent open-minded Bayesian can assign to these newly formulated hypotheses, it can choose to keep assigning a value $P_t(H_{N(t)} \mid \Theta_{N(t)+1}) \geq 1/10 + \varepsilon$ for positive $\varepsilon$. In that case there will be infinitely many stages $t$ at which the one-step predictive probability*

$$\begin{aligned} P_t(0 \mid \Theta_{N(t)+1}) &= \sum_{H \in \Theta_{N(t)+1}} P_t(H \mid \Theta_{N(t)+1}) \cdot H(0 \mid E^t) \\ &> \left(\frac{1}{10} + \varepsilon\right) \cdot H_{N(t)}(0 \mid E^t) \\ &= \frac{1}{10} + \varepsilon, \end{aligned}$$

*blocking convergence to the correct predictive probability $H^*(0 \mid \cdot) = 1/10$.*

This example can be adapted at will to show that for any true $H^*$ there are hypothesis streams and posterior assignments that block convergence. The essential trait is that the newly formulated hypotheses receive—keep receiving—too much posterior. This leads us to an obvious diagnosis: the silent open-minded Bayesian is allowed too much freedom in assigning posteriors to newly formulated hypotheses.

*3.3.2. The silent open-minded version: retroactive variant.* Following up on the previous diagnosis, one way in which it might *seem* we can constrain the freedom of the open-minded Bayesian is to have the posterior be informed by the old evidence. This is the retroactive variant of the silent open-minded Bayesian, Section 2.5 above;

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

but as we explained there already, there is no formal difference between the two versions. That is, barring hypotheses with likelihood 0, any choice of posterior can be modeled as a retroactive choice of prior. This means that any counterexample to the silent open-minded version also yields a counterexample to the retroactive variant, including Example 3.5.

FACT 3. *The retroactive variant of the silent open-minded Bayesian is not guaranteed to weakly merge with the truth whenever the truth is formulated.*

EXAMPLE 3.6. *The calculations now do depend on the likelihoods of all hypotheses on the past data, something that was not specified in Example 3.5. The most straightforward circumstance is where the new hypothesis's likelihood on* $E^t$ *actually equals the agent's probability of* $E^t$ *conditional on* $\Theta_N$,

$$H_N(E^t) = P_0(E^t \mid \Theta_N), \tag{12}$$

*in which case a prior assignment* $P_0(H_N \mid \Theta_{N+1}) := p$ *translates into a posterior* $P_t(H_N \mid \Theta_{N+1}) = p$. *In that case, a prior choice of* $p \geq 1/10 + \varepsilon$ *recovers the previous example. If the new hypothesis's likelihood on the past data is lower than* $P_0(E^t \mid \Theta_N)$, *the prior must be set higher to retrieve the same posterior. As an illustration, if* $H_N(E^t) = 1/3 \cdot P_0(E^t \mid \Theta_N)$, *then a posterior* $p_t > 1/10$ *requires a choice of prior* $p_0 > 1/4$.

*Arguably, however, the more plausible circumstance is for newly proposed hypotheses to have higher likelihood than the earlier hypotheses. Plausibly, new hypotheses, formulated after the past data are in, rather overfit the data: in the most extreme case, actually have a likelihood* 1. *In that case, of course, the same posterior* $p_t$ *requires a smaller prior* $p_0$. *To illustrate again, let the data be generated by* $H_{\theta^*}$ *with* $\theta^* = 0.9$, *so that* $P_0(E^t \mid \Theta_N)$, *with high probability, will not exceed* $H_{\theta^*}$*'s likelihood on the past data* $E^t$, *which for typical data is about* $0.9^{0.9t} \cdot 0.1^{0.1t}$. *In that case, for a new hypothesis* $H_N$ *with* $H_N(E^t) = 1$, *the same posterior only requires an exponentially smaller prior: already for* $t = 10$, *for instance, it suffices for posterior* $p_t > 1/10$ *to set* $p_0 > 1/200$.

The arguably most natural circumstance of new hypotheses that overfit is thus also the most difficult case for our purposes. An extremely modest choice of prior here already suffices for a substantial posterior, and the threat to truth-convergence is precisely such substantial posterior assignments to new and false hypotheses.

One can defend the retroactive approach on the grounds that it accommodates how old evidence confirms new theories ([38], p. 1244f); or one can disown it on the grounds that it involves a "double counting" of the old evidence, since the hypothesis and presumably its prior was already formulated in response to the evidence (cf. [9], 132f). We only point out here that for the above reason of overfitting hypotheses, a retroactive procedure appears more challenging for the aim of truth-convergence. Of course, in the silent version, this cannot make an *essential* difference: both variants are formally equivalent, and the challenge above is limited to a moderate choice of prior in the retroactive variant that does not correspond to a moderate choice of posterior in the original variant. But our analysis below reveals that in the hybrid case, the difference between prior and posterior assignments will matter for the guarantee of truth-convergence.

***3.4. The hybrid open-minded Bayesian.*** The diagnosis from the previous section was clear: the (retroactive) silent open-minded Bayesian is allowed too much freedom in assigning posteriors (priors) to newly formulated hypotheses. Given this diagnosis,

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

one might expect the hybrid version to do better. After all, here there is an explicit constraint on priors: there is only so much the agent can shave off from the catch-all!

This is so, recall, because we interpret the catch-all's prior as at least having some determinate value. This does not exclude that it is "a number extremely close to unity," but it does exclude a conception on which it is some indeterminate value arbitrarily close to 1, perhaps made precise as "unity minus an infinitesimal" ([38], p. 1244). When it comes to truth-convergence, such a conception renders the hybrid version on a par with the silent version: both put no constraints on the choice of prior (posterior), wherefore convergence cannot be guaranteed.[6]

We thus assume, more specifically, that the hybrid version features a certain limited reservoir of prior probability from which the probability for new hypotheses must be taken. We can think of this constraint as simply that, a constraint; we are not committed to understanding this constraint in terms of a catch-all. Nevertheless, we see it as a conceptual plus that it *can* be understood in this way, and this carries over to our own proposal in Section 4.

*3.4.1. Failure of truth-convergence.* Unfortunately, the constraint introduced in the hybrid version does not suffice: we can even produce a scenario where convergence to the true predictive probabilities is *guaranteed to fail*. This scenario again exploits the possibility of a stream of overfitting hypotheses, that despite the constraint on new prior assignments still keep taking up too much posterior. More precisely, on every possible outcome stream we repeat the following: wait while all current probabilistic hypotheses get lower and lower likelihood on the unfolding sequence of outcomes, until the difference with the maximal likelihood of a new overfitting hypothesis is large enough for such a new hypothesis to have a sufficient impact, despite its necessarily constrained prior, on the agent's predictive probabilities.

PROPOSITION 3.7. *The hybrid open-minded Bayesian is not guaranteed to weakly merge with the truth whenever the truth is formulated.*

EXAMPLE 3.8. *Suppose that the true hypothesis is the Bernoulli* $H^* = H_{\theta^*}$ *with* $\theta^* = 1/2$, *and that this hypothesis is indeed formulated at a point in time* $t_0$. *Thus* $H^*$ *is assigned some unconditional prior value* $p^* =: P_0(H^*)$, *leaving the catch-all* $\overline{\Theta}_{N_0+1}$ *with some unconditional prior* $\tau_{N_0+1} = \tau_{N_0} - p^*$.

*Consider a history with* $t_0 < t_1 < t_2 < \cdots$ *infinitely many later points in time at which a new hypothesis is formulated. The vocal open-minded Bayesian is restricted by the prior held by the catch-all in how much prior it can shave off and assign to these new hypotheses; but it can choose to assign each* $H_{N_i}$ *an unconditional prior*

$$P_0(H_{N_i}) = 2^{-i} \cdot \tau_{N_0+1}, \tag{13}$$

*since* $\sum_{i=1}^{\infty} 2^{-i} \cdot \tau_{N_0+1} = \tau_{N_0+1}$.

*Now consider such a history where the newly proposed hypotheses all maximally overfit the past data at their time of formulation, i.e.,* $H_{N_i}(E^{t_i}) = 1$ *for each* $i$, *and then make some biased one-step prediction* $H_{N_i}(0 \mid E^{t_i}) = p_i$, *with* $|p_i - 1/2| > \varepsilon$ *for some fixed* $\varepsilon > 0$.

[6] Wenmackers and Romeijn (ibid.) evoke Earman's worry that the procedure of shaving-off from the catch-all "leads to the assignment of ever smaller initial probabilities to successive waves of new theories until a point is reached where the new theory has such a low initial probability as to stand not much of a fighting chance" ([9], 196). On our analysis, the danger is rather that new theories keep amassing *too much* probability.

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

*Suppose, further, that all hypotheses formulated before the true hypothesis, and all the new hypotheses after their formulation, issue one-step predictive probabilities that are bounded away from* 1: *there is some* $\delta > 0$ *such that all one-step predictive probabilities are smaller than* $1 - \delta$. *The idea is that, whatever the subsequent data, the hypotheses in play will at each point in time leak some of their likelihood, so that, when a new overfitting hypotheses* $H_{N_i}$ *comes in, after the stretch of time between* $t_{i-1}$ *and* $t_i$ *has been large enough, its relative likelihood is so large that its biased prediction will sufficiently distort the overall predictive probability.*

*Specifically, fix some* $\varepsilon' < \varepsilon$, *and let*

$$r = \frac{\frac{1}{2} + \varepsilon'}{\frac{1}{2} + \varepsilon}, \tag{14}$$

*which itself lies in the interval* $\left(\frac{1}{2}, 1\right)$. *Now if at each* $t_i$ *we have*

$$P_{t_i}(H_{N_i} \mid \Theta_{N_i+1}) > r, \tag{15}$$

*then we have for* $E$ *with* $H_{N_i}(E \mid E^{t_i}) > \frac{1}{2} + \varepsilon$ *that*

$$\begin{aligned} P_{t_i}(E \mid \Theta_{N_i+1}) &= \sum_{H \in \Theta_{N_i+1}} P_{t_i}(H \mid \Theta_{N_i+1}) \cdot H(E \mid E^{t_i}) \\ &> P_{t_i}(H_{N_i} \mid \Theta_{N_i+1}) \cdot H_{N_i}(E \mid E^{t_i}) \\ &> \frac{\frac{1}{2} + \varepsilon'}{\frac{1}{2} + \varepsilon} \cdot \left(\frac{1}{2} + \varepsilon\right) \\ &= \frac{1}{2} + \varepsilon', \end{aligned}$$

*blocking convergence.*

*As worked out in Appendix* 6.2, *inequality* (15) *is guaranteed if each*

$$t_i - t_{i-1} > \frac{-\log(1-r) - (-\log r) + i - \log \tau_{N_0+1}}{-\log(1-\delta)}. \tag{16}$$

*To break* (16) *down a little, note that if* $\varepsilon$ *is reasonably large, and* $\varepsilon'$ *chosen very small, then* $r$ *is relatively close to* $1/2$ *and has a minor influence on the bound. For instance, if* $r < 2/3$, *which would follow from* $\varepsilon > 1/4$ *and* $\varepsilon' \approx 0$, *then* $-\log(1-r) - (-\log r) < 1$, *so that* (17) *is already implied by*

$$t_i - t_{i-1} > \frac{1 + i - \log \tau_{N_0+1}}{-\log(1-\delta)}. \tag{17}$$

*Furthermore, we have* $\delta = 1/2$ *and* (17) *reduces to*

$$t_i - t_{i-1} > 1 + i - \log \tau_{N_0+1} \tag{18}$$

*in the extreme case where all hypotheses except* $H_{N_i}$ *after* $t_{i-1}$ *always give predictive probabilities* $1/2$.

*3.4.2. Discussion.* The failure of truth-convergence of the hybrid open-minded agent may strike one as surprising. It is, after all, characteristic of the hybrid procedure that the true hypothesis, once formulated, holds an explicitly assigned share $p^* > 0$ of the absolute prior. As soon as the true hypothesis is formulated, the unconditional agent

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

function $P_0$ holds a grain $p^*$ of this truth, *no matter what hypotheses with what priors are still added later*. This carries over to the retroactive prior measures conditional on any hypothesis set after the truth is formulated: $P_0(H^* \mid \Theta_N) \geq p^*$ for all hypothesis sets $\Theta_N$ after the formulation of $H^*$. But does this not suggest that the agent function holds a grain of the truth, and was this not already enough for strong truth-merger?

A complete answer to what is wrong with this intuition requires us to make perfectly precise the desideratum of an open-minded agent's truth-convergence. We will here first briefly make the above intuition precise in a way that is *faulty*, but that allows us to highlight the challenges we face in formalizing our desideratum of an open-minded agent's truth-convergence. In the next section we proceed to meet these challenges and formalize our desideratum, to subsequently propose a version of an open-minded Bayesian that does satisfy a version of truth-convergence.

Thus let us for a moment consider the retroactive measure $P_0(\cdot \mid \Theta_\infty)$, induced by the actually generated hypotheses and prior assignments *in the limit*. This measure must also hold a grain $p^*$ of the truth. What, exactly, is unsatisfying about proclaiming truth-convergence of the open-minded agent, from the fact that we can always derive, with Corollary 3.3, strong truth-merger of *this* measure?

The straightforward answer is that this measure's merger must be unsatisfying because, as we already know from Example 3.8, it can go together with a *guaranteed failure of convergence* of the agent's one-step predictive probabilities. But how can this be? Here it is important to note that, in Example 3.8, the hypothesis stream emphatically depends on the actually generated data stream. While the agent function $P_0(\cdot \mid \Theta_\infty)$ induced by this particular data and hence hypotheses stream can be shown to merge with $H^*$ (as it contains a grain of $H^*$), this is still consistent with it *failing to converge on the actual data stream that induced it*. (The latter is consistent with a.s. truth-convergence, because, in our example, any particular outcome stream that is actually generated is an $H^*$-probability-0 event.)

This provides an illustration of the two challenges we mentioned in Section 3.2. First, since we have an hypothesis stream as a moving part, we have to be very careful with the interpretation of probability-1 statements on the data space. The retroactive agent function $P_0(\cdot \mid \Theta_\infty)$ was only put in place, so to speak, after already fixing the actually generated data stream, and the strong merger only derived after that. In contrast, intuitively, the "almost sure" should range over the possible data *and all that depends on it*, including the possible hypotheses (hence possible shapes of the agent function) that are formulated in response to it. The challenge is to attain a formal a.s. convergence that is also still meaningful in this sense. This is intertwined with the second challenge, which is to make precise which agent function we actually seek convergence for. The obvious diagnosis is that the functions $P_0(\cdot \mid \Theta_\infty)$, having this "after the fact" quality of being dependent on a particular data and hence hypothesis stream, and indeed of then having available this hypothesis set *from the start*, are not what we are after.

We now proceed to look for an answer to these two challenges, towards reclaiming a property of truth-convergence.

**§4. Forward-looking open-minded truth-convergence.** We further analyze the goal of truth-convergence, introducing the notion of an hypothesis and posterior scheme and that of a completed agent measure (Section 4.1). We then propose a *forward-looking* open-minded Bayesian, the completed agent measure of which *does* retain a grain of the truth, from which the agent's weak merger follows. We first propose a

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

proto-variant of this version, which is a variant of the silent open-minded Bayesian with a limited posterior reservoir (Section 4.2), before we introduce the final version, a variant of the hybrid open-minded Bayesian with a restriction on new hypotheses' likelihoods (Section 4.3).

### *4.1. Towards regaining truth-convergence.*

*4.1.1. The hypothesis scheme.* We start with the first challenge in drawing up the desired convergence statement: how should we think about the "almost surely"? In the following, we suppose for simplicity of presentation that the agent possesses the true hypothesis $H^*$ from the start, $H^* \in \Theta_0$.[7]

We first observe that it is impossible to derive a statement of the following form.

(i) For every $H^*$, there is an $H^*$-measure-1 class of infinite output streams on which the open-minded agent converges to $H^*$, independent of the stream of newly formulated hypotheses.

Already in the case of the standard Bayesian agent, the $H^*$-measure-1 class of output streams on which the agent converges cannot generally be independent of the other elements in the agent's hypothesis class. Consider for the true $H^*$ again the Bernoulli-$1/2$ measure: it is not hard to see that for each possible infinite outcome stream, there exist hypothesis sets that contain $H^*$ yet are such that the agent does not converge on *this* outcome stream. As an extreme case, the agent will not converge on outcome stream $E^\omega$ if the hypothesis set contains an hypothesis that assigns probability 1 to this exact sequence $E^\omega$: the agent will converge, not on the true predictive probabilities $1/2$, but on predictive probabilities 1 for the correct next outcomes. This example concerns the initial hypothesis set of a standard (or indeed open-minded) agent, but easily transfers to the streams of newly formulated hypotheses given to any plausible version of an open-minded agent.[8] Thus a statement of form (i) is too strong.

This leads us to the following statement, where we have shifted the quantifiers to allow the exact measure-1 class to depend on the hypothesis stream.

(ii) For every $H^*$, every hypothesis stream, there is an $H^*$-measure-1 class of infinite outcome streams on which the open-minded agent converges to $H^*$.

In order to demonstrate a statement of the form (ii), we must prove, for any given hypothesis stream, convergence on the presupposition of *this* stream. Formally, we conceive of $\Theta_{N(\cdot)}$ as a function that maps each time $t$ to an hypothesis set $\Theta$. Of course, this function must also return hypothesis sets in such a way that they actually correspond to some possible open-minded agent. For instance, for each $t$ there can be at most one hypothesis in $\Theta_{N(t+1)} \setminus \Theta_{N(t)}$.

[7] For the general case where the truth is formulated after some finite time $t$, or more specifically, after some finite sequence $E^t$, mentions of 'an $H^*$-measure-1 class of infinite outcome streams' should be replaced by 'an $H^*(\cdot \mid E^t)$-measure-1 class of infinite outcome streams extending $E^t$', and the 'stream (scheme) of newly formulated hypotheses' by the 'stream (scheme) of newly formulated hypotheses after $E^t$'.

[8] We only need to assume that the agent's posteriors will indeed converge on the predictions of hypotheses that perform *perfectly*, which is a minimal condition for a version that will in fact have the capacity to converge to the truth.

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

There is a clear sense, however, in which a statement of form (ii) is too weak. The main challenge for establishing truth-convergence is, recall Example 3.8, the possibility of overfitting hypotheses *in reaction to each possible outcome stream*. In light of such scenarios, presupposing a particular hypothesis stream, irrespective of the generated data, is obviously unsatisfying.

But we may just as well assume that the generation of hypotheses is given by a function that links hypothesis sets, not simply to the possible points in time, but to all *possible finite outcome sequences*. That is, we presuppose some *data-dependent* (what we shall call) *scheme* for generating hypotheses, or simply *hypothesis scheme*, that is a function $\Theta_{(.)}$ that maps each finite data sequence $E^t$ to an hypothesis set $\Theta_{E^t}$. Again, this function must also be constrained by the open-minded agent's specification.

This then leads us to aim for a convergence statement of the following form.

(iii) For every $H^*$, every hypothesis scheme, there is an $H^*$-measure-1 class of infinite outcome streams on which the open-minded agent converges to $H^*$.

Note that the assumption of a particular $H^*$ in conjunction with an hypothesis scheme comes down to treating hypothesis streams as *random* quantities, as they are given by a function on the outcome streams governed by probability measure $H^*$. One could take this further and consider for the true measure more elaborate probabilistic models that also directly range over the class of possible hypothesis streams. We do not go this way here: we stick here to a true measure $H^*$ that is a function over outcome sequences only, and work towards a convergence statement where the $H^*$ measure-1 class can depend on the hypothesis scheme. Of course, there is more to say about the conceptual status of a convergence statement of the form (iii), and we will say a bit more below.

We first observe, however, that there is still something left implicit in statement (iii). This is the agent's actual choice of posteriors (or, depending on the version, retroactive choice of priors resulting in posteriors) for the incoming hypotheses.

*4.1.2. The posterior scheme.* But given a particular hypothesis scheme, perhaps we could always derive convergence for a particular $H^*$-measure-1 class of outcome streams, that *is* independent of the exact (positive) posterior values the agent chooses to assign to these incoming hypotheses?

Unfortunately, this is again not attainable in general. Again we indeed already have for the standard Bayesian agent that a different choice of prior distribution over the exact same hypothesis set (more exactly, a different *regular* prior distribution that assigns each element positive probability) can result in a different $H^*$-measure-1 class of outcome sequences on which it converges to $H^*$. In fact, we can show that there are single hypothesis sets such that for *every* individual stream we can tweak the priors in such a way that convergence fails on *this* stream.

PROPOSITION 4.9. *There exist countable hypothesis sets* $\Theta$ *and hypotheses* $H^* \in \Theta$ *such that for every infinite outcome stream* $E^\omega$, *there is a regular prior distribution* $P$ *over* $\Theta$ *such that the Bayesian agent* $P$*'s predictive probabilities do not converge to* $H^*$ *on* $E^\omega$.

*Proof.* See Appendix 6.3. □

This result pertains to the initial hypothesis set of a standard (or indeed open-minded) agent, but the initial set is already part of an open-minded agent's hypothesis

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

scheme, and the result could again readily be modified to pertain to the posterior assignments to a scheme's newly formulated hypotheses. Thus the result implies that we must allow the measure-1 class to also depend on the *posterior scheme*, that specifies what numerical posterior values are assigned to each (incoming) hypothesis. Formally, the combination of the hypothesis and the posterior scheme is now codified in a function $P_{(\cdot)}$ that maps each finite data sequence $E^t$ to a posterior distribution $P_{E^t}$ over the hypothesis set $\Theta_{E^t}$. Again, this function must also return distributions that actually correspond to some possible open-minded agent; that is to say, these distributions must be consistent with the specifications of the version of the open-minded agent in question. For instance, in case of the hybrid agent, the distribution $P_{E^t}$ is the distribution $P_t(\cdot \mid \Theta_N)$ after having observed $E^t$ and with $\Theta_N = \Theta_{E^t}$. By the specification of the hybrid agent, this distribution $P_t(\cdot \mid \Theta_N) = P_0(\cdot \mid E^t, \Theta_N)$ is derived from some prior distribution $P_0$ over $\Theta_N$. This latter distribution must cohere with the priors $P_0(\cdot \mid \Theta_{N'})$ for earlier and later hypothesis sets $\Theta_{N'}$, which likewise constrain the distributions $P_{E^s}(\cdot) = P_s(\cdot \mid \Theta_{N'})$ for $E^s$ that extend or are extended by $E^t$. Whenever we invoke hypothesis and posterior schemes in the following, we implicitly limit our attention to schemes that actually correspond to open-minded agents of the version we are then considering.[9]

This then leads us, finally, to aim for a convergence statement of the following form.

(iv) For every $H^*$, every hypothesis and posterior scheme, there is an $H^*$-measure-1 class of infinite outcome streams on which the open-minded agent converges to $H^*$.

Having thus derived the formal structure of the strongest statement we can hope for, let us expand a little bit on its conceptual status. One possible interpretation is that this statement corresponds to an assumption that prior to the inquiry, both the future hypotheses and the posteriors that will be assigned to them are, albeit still dependent on the random data and unknown the agent, already fixed. There is at least a superficial tension between such an interpretation and a crucial motivation for investigating open-minded agents, namely that hypotheses and their priors are not fixed in advance, and the agent has full freedom to choose these on the go. How problematic this is, would then conceivably depend on one's view on the external process where the hypotheses and posteriors come from: is this in the end some mechanical procedure, or is this some process of creative and fundamentally opaque scientific discovery? On the other hand, we think it is actually not so clear that the mathematical structure of a statement of form (iv), "fix arbitary $x$, we now show...," commits one to a conceptual view of the kind, "assuming that $x$ is fixed in advance, we have that...," let alone what it exactly means for an hypothesis scheme to be (unknown to the agent but) determined in advance. These are philosophically murky waters, and we will here limit ourselves to noting that mathematically, this is the best we can aim for. Indeed, if already for the standard

[9] Some care is required in deriving relations between the functions $P_{E^t}(\cdot \mid \Theta_{E^t})$ from the agent specifications, which also involves matching the original notation for agent functions ("$P_t(\cdot \mid \Theta_N)$") with the $P_{E^t}(\cdot \mid \Theta_{E^t})$. The former notation leaves implicit what exactly are the past data that have resulted in the posteriors and hypothesis sets, which becomes especially risky when analyzing retroactive assignments (what future hypothesis set and posteriors is $P_0(\cdot \mid \Theta_N)$ actually reconstructed from?). This will mostly matter for the proofs to follow: see Appendix 6.1 on notation used there for details.

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

Bayesian agent the precise measure-1 class must depend on the other hypotheses and exact priors, it is only natural to aim for the analogous statement for the open-minded agent—in general. This does not exclude the possibility of deriving statements of form (i) with certain restrictions on the possible hypotheses, say a restriction of effective computability. But this lies out of the scope of the current paper.

With this conceptual provision, we are now clear on the nature of the "almost surely" qualification. In fact, we have also already touched on the second challenge: what, exactly, is the agent function that we seek convergence for? We will now proceed to make this perfectly precise.

*4.1.3. The completed agent measure.* Given an hypothesis and a posterior scheme, an open-minded Bayesian's probability assignments after each possible finite outcome sequence are fully determined. For all finite $E^t$, the agent's assignment to any event $A$ is fixed and given by

$$P_{E^t}(A) = P_{E^t}(A \mid \Theta_{E^t}). \tag{19}$$

The corresponding convergence statement of form (iv), for *strong* merger, is that for each hypothesis and posterior scheme, we have for an $H^*$-measure-1 class of infinite outcome sequences that

$$\sup_{A \in \mathfrak{F}} \left| P_{E^t}(A) - H^*(A \mid E^t) \right| \xrightarrow{t \to \infty} 0. \tag{20}$$

Here we still adhered to the simplifying assumption made at the beginning of Section 4.1.1, that the truth $H^*$ is contained in the initial hypothesis class. The general case is covered by adding the formulation of $H^*$ on the outcome stream as a condition for the convergence. That is, for an $H^*$-measure-1 class of infinite outcome sequences,

$$H^* \text{ is formulated} \implies \sup_{A \in \mathfrak{F}} \left| P_{E^t}(A) - H^*(A \mid E^t) \right| \xrightarrow{t \to \infty} 0. \tag{21}$$

For *weak* merger, this comes down to

$$H^* \text{ is formulated} \implies \sup_{E_{t+1} \in \{0,1\}} \left| P_{E^t}(E_{t+1}) - H^*(E_{t+1} \mid E^t) \right| \xrightarrow{t \to \infty} 0. \tag{22}$$

A circumstance that makes convergence of the terms (19) hard to analyze is that, even under the assumption of a given hypothesis and posterior scheme, *they may not correspond to a single probability measure*. That is to say, the assignments $P_{E^t}(\cdot)$ cannot in general be reconstructed as the conditional probabilities of a particular measure: there need not be a single measure $P$ such that $P(\cdot \mid E^t) = P_{E^t}(\cdot)$ for each $E^t$. This stems from the fact that an open-minded agent's assignments can be dynamically incoherent, in the sense that for finite sequences $E^{t_1}, E^{t_2}$, the second extending the first,

$$P_{E^{t_1}}(A \mid E^{t_2}) \neq P_{E^{t_2}}(A). \tag{23}$$

In words, the agent's assignment to event $A$ at time $t_1$, conditional on the extended outcome sequence $E^{t_2}$, may not equal the agent's assignment to $A$ at time $t_2$, after having in fact seen $E^{t_2}$. To make this slightly more concrete, consider again the hybrid open-minded agent. From its specification, there is some prior distribution $P_0$ such that $P_{E^{t_1}}(A \mid E^{t_2}) = P_0(A \mid E^{t_2}, \Theta_{E^{t_1}})$ and $P_{E^{t_2}}(A) = P_0(A \mid E^{t_2}, \Theta_{E^{t_2}})$. But there is

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

no reason why the terms $P_0(A \mid \Theta_{E^t1})$ and $P_0(A \mid \Theta_{E^t2})$, conditional on different hypotheses, should be equal.

Nevertheless, the agent's *one-step* predictive probabilities, given a particular hypothesis and posterior scheme, *do* induce a coherent set of probability assignments. The predictive probabilities $P_{E^t}(E_{t+1})$ induce a probability assignment $P^\infty$ on all finite evidence sequences, by

$$P^\infty(E^t) := \prod_{i=0}^{t-1} P_{E^i}(E_{i+1}), \tag{24}$$

and this induces a measure on all outcome streams. We will call this measure $P^\infty$ the *completed agent measure*.

If we are able to show that, for any given hypothesis and posterior scheme, this measure retains a grain of the truth $H^*$, then a statement of form (iv), for *strong* merger, follows from Corollary 3.3. That is, for any given hypothesis and posterior scheme, we can conclude that for an $H^*$-measure-1 class of outcome streams,

$$H^* \text{ is formulated} \Longrightarrow \sup_{A\in\mathfrak{F}} \left|P^\infty(A \mid E^t) - H^*(A \mid E^t)\right| \xrightarrow{t\to\infty} 0. \tag{25}$$

However, this statement concerns the completed agent measure $P^\infty$, and not the open-minded agent's actual assignments at each time, given by (19). These assignments $P^\infty(A \mid E^t)$ and $P_{E^t}(A)$ may not coincide. The potential disagreement lies in the fact that $P^\infty(A \mid E^t)$ is already influenced by what *future* hypotheses, formulated after $E^t$, say about $A$; whereas $P_{E^t}(A)$ only depends on the hypothesis set $\Theta_{E^t}$.

Still, we do have by definition that these functions coincide on the one-step predictive probabilities. We have that $P^\infty(E_{t+1} \mid E^t) = P_{E^t}(E_{t+1})$ for each outcome sequence $E^t$ and single outcome $E_{t+1}$, so that convergence statement (25) does imply convergence statement (22).[10]

Thus, if we can show, for any given hypothesis and posterior scheme, that the open-minded agent's completed agent measure holds a grain of the truth, then we can derive a convergence statement of form (iv) for *weak* merger of the agent functions. Consequently, in the following, we will work towards ensuring this property, that the completed agent measure holds a grain of the truth, whenever the truth is formulated.

*4.1.4. The failure of holding a truth-grain.* Consider again the hybrid open-minded agent. Connecting back to the discussion of Section 3.4.2, it might seem that the completed agent measure should hold a grain of the truth as soon as for every single $E^t$, the retroactive prior function $P_0(\cdot \mid \Theta_{E^t})$ holds at least a grain $p^*$ of $H^*$; that is, whenever all these $P_0(\cdot \mid \Theta_{E^t})$ *uniformly* retain at least the same grain of the truth. This, however, is *not* so.

That this cannot be so is again already implied by Example 3.8. This example in fact features a (partially specified) hypothesis and posterior scheme for overfitting hypothesis generation, where every $P_0(\cdot \mid \Theta_{E^t})$ for $t \geq t^*$ holds at least a grain $p^*$ of the truth. Yet we saw that the agent (the completed agent measure) in that example

[10] In fact, for any $t$, measures $P^\infty(\cdot \mid E^t)$ and $P_{E^t}$ coincide up to the smallest time ahead at which a new hypothesis will be formulated; though this only implies weak convergence of the latter for $d > 1$ if this time horizon will eventually always be at least $d$.

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

fails to merge with $H^*$, which by the contraposition of Corollary 3.3 entails that the completed agent measure cannot hold a grain of $H^*$.

PROPOSITION 4.10. *For the hybrid open-minded Bayesian, there are hypothesis schemes with $H^* \in \Theta_0$ such that nevertheless the completed agent measure fails to hold a grain of the truth: there is no $a \in (0, 1)$ with $P^\infty(E^t) \geq a \cdot H^*(E^t)$ for all $E^t$.*

*Proof.* Such a scheme is given by Example 3.8: see Appendix 6.4 for details. □

What, intuitively, explains this fact, that each $P_0(\cdot \mid \Theta_{E^t})$ can uniformly hold a grain of the truth, yet $P^\infty$ does not? The difference between each of the former functions and $P^\infty$ is that in the latter, overfitting hypotheses are not represented in the predictive probabilities issued by the agent until this hypothesis actually comes in. But by definition these overfitting hypotheses have high likelihood (and thus issue high predictive probabilities) on these initial segments; so taking them out will deflate the agents' predictive probabilities on these initial segments. The counterexample shows that this effect can be so strong that it destroys the grain of the truth.

In our proposal of a *forward-looking* open-minded Bayesian, that we turn to now, we focus on making sure that the completed agent measure does retain a grain of the truth, whenever the truth is formulated, in order to derive a guarantee of truth-convergence.

***4.2. The forward-looking open-minded Bayesian, proto-version.*** We first consider a version of an open-minded Bayesian, a proto-version of the forward-looking open-minded Bayesian that we propose in Section 4.3, that rests on the following simple idea. Instead of a limited reservoir of probability for assigning *priors* to new hypotheses, the agent has a limited reservoir of *posterior* mass to assign to new hypotheses.

*4.2.1. Specification.* The forward-looking open-minded agent, in this proto-version, is like the silent open-minded agent, in that we do not stipulate a catch-all or a limited absolute reservoir of prior probability. However, we do stipulate a limited absolute reservoir of *posterior* probability: unlike the silent open-minded Bayesian, that can assign any posterior to a new hypothesis, the agent must shave off a new posterior from this reservoir, thereby shrinking the reservoir for posterior assignments to future new hypotheses. We assume that the starting reservoir holds a certain real-valued mass $d > 0$ (we do not need to assume that this mass is bounded by 1). In addition, as a minimal restriction that facilitates the proof of truth-convergence, we assume that there is a constant $c < 1$ such that agent is not allowed to assign a posterior greater than $c$ to any single new hypothesis.

In summary, the **proto-version** of the **forward-looking open-minded Bayesian** proceeds as follows.

($t = 0$) *$N$ explicit hypotheses.* As in the silent version, each explicit hypothesis $H_i$ in $\Theta_N$ receives a prior $P_0(H_i \mid \Theta_N) > 0$ conditional on $\Theta_N$, such that $\sum_{i<N} P_0(H_i \mid \Theta_N) = 1$. In addition, there is assumed a reservoir $\tau_N = d > 0$ of posterior probability, and a maximal one-time probability $c < 1$.

($t > 0$) *Evidence $E^t$.* Updating proceeds in the usual way, conditional on the current hypothesis set $\Theta_N$.

($t > 0$) *New hypothesis $H_N$.* As in the silent version, when a new hypothesis $H_N$ is formulated, extending the hypothesis set to $\Theta_{N+1} = \Theta_N \cup \{H_N\}$, the posterior $P_t(H_N \mid \Theta_{N+1})$ is directly set to a value $p_N$; but now this value $p_N \leq c$ must be obtained from

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

decomposing the posterior reservoir $\tau_N$ into $p_N$ and a remainder $\tau_{N+1} = \tau_N - p_N$ that is the new posterior reservoir.

*4.2.2. Verification.* The forward-looking open-minded Bayesian's constraints in attributing posterior mass to newly formulated hypotheses rules out a scenario like Example 3.8, where constrained prior assignments still lead to high posterior values. As a matter of fact, the restriction on posterior values results in a completed agent measure that *does* retain a grain of the truth, whenever it is proposed.

THEOREM 4.11. *For the proto-version of the forward-looking open-minded Bayesian, for any hypothesis and posterior scheme, the completed agent measure conditional on any* $E^t$ *with* $H^* \in \Theta_{E^t}$ *holds a grain of* $H^*$.

*Proof.* See Appendix 6.5. □

COROLLARY 4.12. *For the proto-version of the forward-looking open-minded Bayesian, for any hypothesis and posterior scheme, we have that* $H^*$*-a.s.*

$$H^* \text{ is formulated} \Longrightarrow \sup_{E_{t+1} \in \{0,1\}} \left| P_{E^t}(E_{t+1}) - H^*(E_{t+1} \mid E^t) \right| \xrightarrow{t \to \infty} 0.$$

*4.2.3. Discussion.* As mentioned, this proto-version of a forward-looking Bayesian is a constrained version of the silent open-minded Bayesian. More precisely, it is a constrained version, not of the retroactive, but of the *standard* variant of the silent Bayesian. The posteriors of new hypotheses are chosen directly; and however this is done (within the constraint of the posterior reservoir), it is not required to be (not part of the agent's specification to be) an explicit calculation of the posterior from a chosen prior and the hypothesis's likelihood on the past outcome sequence.

Again, the choice of posterior *can* always proceed like this: formally, any choice of posterior corresponds, via the likelihood on the past data, to a choice of prior. But the *constraint* on the posteriors does not translate into a simple constraint on the priors, depending as it does on the contingent fact of the actually formulated hypotheses' likelihoods, and so a *retroactive* variant of the forward-looking Bayesian does not appear a natural option—as, of course, its name is intended to suggest.

That said, the idea of an absolute reservoir of posterior probability is not a terribly natural conception. Unlike the idea of an absolute reservoir of prior probability, it cannot be coupled to a conception of a prior assignment to a catch-all hypothesis, from which new hypotheses may be shaven off. Perhaps the best way to understand this is simply as a pragmatic device, that is easy to understand and does the job of regaining the guarantee of truth-convergence.

However, we think there is yet a conceptually more satisfying option, that is formally very similar to the current version but that has a more natural interpretation. In fact, this version, our actual forward-looking Bayesian, *does* regain the idea of shaving prior mass from a catch-all, while still looking forward.

***4.3. The forward-looking open-minded Bayesian.*** An alternative way of defusing the threat of extreme posteriors of incoming hypotheses is to place restrictions, not directly on the posteriors, but on the *likelihoods* of new hypotheses. Our proposal is to introduce the stipulation that new hypotheses have some default likelihood on past outcomes.

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

We will focus on an idea that we borrowed from the theory of competitive online learning[11], and that has important technical and conceptual advantages. This idea is to identify the likelihood of new hypotheses on past data with the *agent's* probability assignment to this data, induced from its past predictive probabilities. That is, a new hypothesis $H_N$'s likelihood $H_N(E^t)$ on the data sequence $E^t$ generated by its time $t$ of formulation is set equal to the product $\prod_{s=0}^{t-1} P_0(E_{s+1} \mid E^s, \Theta_{N(s)})$ of predictive probabilities. Note that this is precisely the completed agent measure's assignment $P^\infty(E^t)$.[12]

This is a natural way of modeling that a new hypothesis is only evaluated *after* its formulation; or that with respect to this new hypotheses, the old evidence does not count. The new hypothesis is, to put it differently, at its time of formulation treated in a *neutral* fashion, in that it is supposed to have had the same predictive success on the past data as the agent itself. This also translates in this new hypothesis having, for any chosen prior $P_0(H_N \mid \Theta_{N+1})$, at its time of formulation $t$ a *posterior* $P_0(H_N \mid E^t, \Theta_{N+1})$ that simply *equals the prior*.

Moreover, this allows us to recover the picture of a catch-all, or more precisely, the fixed well of prior probability from which the agent must draw in its assignment to (new) hypotheses. In combination with the restriction on prior assignments that this entails, this version of a forward-looking Bayesian indeed regains truth-convergence.

*4.3.1. Specification.* The forward-looking open-minded Bayesian, in its current version, proceeds exactly as the hybrid-open minded Bayesian, except for the crucial stipulation that each new hypothesis $N_i$ formulated at time $t_i$ satisfies

$$H_{N_i}(E^t) := P^\infty(E^t) \text{ for all } t \leq t_i. \tag{26}$$

In summary, the **forward-looking open-minded Bayesian** proceeds as follows.

$(t = 0)$ *N explicit hypotheses.* As in the hybrid version, each explicit hypothesis $H_i$ in $\Theta_N$ receives a prior $P_0(H_i \mid \Theta_N) > 0$ conditional on $\Theta_N$, such that $\sum_{i<N} P_0(H_i \mid \Theta_N) = 1$; and the catch-all hypothesis $\overline{\Theta}_N = \Theta \setminus \Theta_N$ receives an unconditional prior $P_0(\overline{\Theta}_N) := \tau_N$, so that the unconditional priors of the explicit hypothesis are given by $P_0(H_i) := (1 - \tau_N) \cdot P_0(H_i \mid \Theta_N)$.

$(t > 0)$ *Evidence* $E^t$. Updating proceeds in the usual way, conditional on the current hypothesis set $\Theta_N$.

$(t > 0)$ *New hypothesis* $H_N$. As in the hybrid version, when a new explicit hypothesis $H_N$ is formulated, extending the hypothesis set to $\Theta_{N+1} = \Theta_N \cup \{H_N\}$, the unconditional prior $\tau_N$ of the earlier catch-all is decomposed into a value $p < \tau_N$ for the unconditional prior $P_0(H_N)$ of the new hypothesis and a remainder $\tau_{N+1} = \tau_N - p$ for the unconditional prior $P_0(\overline{\Theta}_{N+1})$ of the new catch-all. The priors conditional on the new hypothesis set are obtained by renormalization, from which the conditional

[11] See [4] for a general account of competitive online learning or prediction with expert advice. The idea that we refer to, first proposed, within the setting of *specialists* [10], by Chernov and Vovk [5], is known as the *specialist* or *abstention trick*; also see [21] and [26]. An instance of this idea also appears in [30].

[12] An alternative way of phrasing—or enforcing—this stipulation is that we do not change hypotheses' past assignments, but only *allow* hypotheses that satisfy this equality.

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

posteriors are obtained by the usual updating on their likelihoods, where the new hypothesis's likelihood $H_N(E^t)$ is stipulated to equal $P^\infty(E^t)$.

*4.3.2. Verification.* Although they differ in their interpretation and also slightly in the precise shape of the formal constraints, the forward-looking Bayesian and its proto-version share the property of a constraint on new posterior assignments. In Appendix 6.5 we give a general proof that for both types of constraints shows that a completed agent measure will hold a grain of the truth, whenever it is formulated, from which weak merger of the agent follows.[13]

THEOREM 4.13. *For the forward-looking open-minded Bayesian, for any hypothesis and posterior scheme, the completed agent measure conditional on any* $E^t$ *with* $H^* \in \Theta_{E^t}$ *holds a grain of* $H^*$.

*Proof.* See Appendix 6.5. □

COROLLARY 4.14. *For the forward-looking open-minded Bayesian, for any hypothesis and posterior scheme, we have that* $H^*$*-a.s.*

$$H^* \text{ is formulated} \Longrightarrow \sup_{E_{t+1}\in\{0,1\}} \left|P_{E^t}(E_{t+1}) - H^*(E_{t+1} \mid E^t)\right| \xrightarrow{t\to\infty} 0.$$

*4.3.3. Beyond weak merger.* Corollary 4.14 states, for the forward-looking agent, and as a consequence of the strong truth-merger of the completed agent measure, the weak truth-merger (with $d = 1$) of the agent measures $P_{E^t}$. The obvious further question is whether we also have strong merger, or at least weak merger for any finite $d$, for the agent measures $P_{E^t}$. We conjecture that already strong merger does hold, but unfortunately we have no proof, and must leave this as an open question.[14]

[13] An alternative proof proceeds by deriving from the abstention stipulation (26) that the forward-looking agent's probability $P^\infty(E^t)$ must *coincide* with the retroactive prior probability $P_0(E^t \mid \Theta_{N_i})$ for every $\Theta_{N_i}$ with $t_{i+1} > t$. The additional stipulation of a fixed amount of prior mass guarantees again that these $P_0(E^t \mid \Theta_{N_i})$ indeed uniformly retain a grain of the truth, so that truth-merger follows. Recall from Section 4.1.4 that the hybrid open-minded Bayesian's completed agent measure can fail to retain a grain of the truth even if every $P_0(\cdot \mid \Theta_{N_i})$ for $i \geq i^*$ uniformly does so: stipulation (26) thus rules out this possibility.

[14] For any infinite $E^\omega$ in the measure-1 class of infinite streams on which we, for given hypothesis and posterior scheme, have strong merger with $H^*$ of the completed agent measure, it might seem that strong truth-merger of the agent functions $P_{E^t}(\cdot \mid \Theta_{E^t})$ on this $E^\omega$ should follow, too: as the posterior reservoir is used up the measures $P^\infty(\cdot \mid E^t)$ and $P_{E^t}(\cdot \mid \Theta_{E^t})$ can differ less and less. However, on any individual $E^\omega$, it is possible that the posterior reservoir is *not* fully used up: this allows for a counterexample, on this particular stream, where the same constant posterior keeps being assigned to new hypotheses on side-branches of $E^\omega$ to force a difference between $P^\infty(\cdot \mid E^t)$ and $P_{E^t}(\cdot \mid \Theta_{E^t})$. Now one could push further and consider the measure-1 class that is the countable intersection of the previous class and, for *every* length $s$, the measure-1 class of streams on which every measure $P_{E^s}(\cdot \mid E^s)$, from that point treated as a standard Bayesian, strongly merges with $H^*$. But even for a stream $E^\omega$ in this class, it is still consistent that the agent measures $P_t(\cdot \mid E^t)$ do not strongly merge with $H^*$ on this particular $E^\omega$: at the same time, such a scenario is now so bizarre that it does not seem feasible to turn it into an actual counterexample, for which this must actually happen with positive probability. This invites the hope for some (martingale) argument that such scenarios must indeed have probability 0.

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

**§5. Conclusion.** We investigated the failure of truth-convergence for Wenmackers and Romeijn's versions of open-minded Bayesianism, and, towards reclaiming this property, proposed a *forward-looking* open-minded Bayesian. The general threat to convergence to the truth is the possibility of new and false hypotheses that keep receiving too much posterior: either by direct assignment or by retroactive calculation from a high likelihood on the past evidence. The proto-version and the final version of our forward-looking Bayesian implement the two respective ways of meeting this threat: by restricting the posteriors, or by restricting the priors and likelihoods.

We think that the final version of our forward-looking agent, which is based on an idea from the theory of competitive online learning, indeed provides an elegant account of how a Bayesian agent should deal with newly formulated hypotheses. The idea of identifying a new hypothesis's likelihood with the agent's probability assignment on the past data is a graceful way of neutralizing the impact of old evidence. Moreover, this idea has the pleasant consequence that the stipulation of a limited reservoir of prior probability (with the associated interpretation of a catch-all hypothesis) is sufficient to guarantee truth-convergence. Unlike the proto-version, that we ourselves feel is mainly a technical device geared towards the aim of truth-convergence, we think the final version makes intuitive sense quite independent of this aim.

There are a number of avenues for further investigation. Firstly, we proved, more precisely, the forward-looking agent's weak truth-merger, or a.s. convergence to the true one-step predictive probabilities. We leave as an open question whether this may be extended to an arbitrary finite-length horizon, or even to strong merger, that includes all tail events. Secondly, a possible lingering doubt is that in our convergence statement the measure-1 class of sequences is dependent on the hypothesis and posterior scheme. This at least suggests an interpretation where the latter are somehow fixed prior to the inquiry, which, one might feel, does not sit well with the original motivation for investigating an open-minded agent. Whether or not this is so, we showed that in general one cannot avoid this dependence, as an analogue in fact already holds in the case of the standard Bayesian. Nevertheless, it might be avoided as further refinements are added to our proposal. Perhaps the main peculiarity about our approach is that in the course of an inquiry hypotheses are not (should not be) introduced haphazardly. There will normally only arise a need for formulating a new hypothesis if some misfit between the data and the current model is observed, which may indeed be regulated via a formal model verification procedure. This raises the question, finally, how (our version of) an open-minded Bayesian inductive logic may be extended beyond just *how* to incorporate externally proposed hypotheses, to also include *when* to accept such new hypotheses, and how this interacts with the guarantee of truth-convergence.

## §6. Appendix: Calculations and proofs.

***6.1. Notation.*** We introduce additional notation for use in the appendices.

For sequences $E^t$ and $E^s$ we write $E^t \preccurlyeq E^s$ if $E^t$ is an initial segment of $E^t$, and $E^t \prec E^s$ if $E^t \preccurlyeq E^s$ and $E^t \neq E^s$. We write $E^t \mid E^s$ if neither $E^t \preccurlyeq E^s$ nor $E^s \preccurlyeq E^t$. For the concatenation of sequences $E^t$ and $E^s$ we write $E^{t+s} = E^t E^s$. For sequences $E^t \preccurlyeq E^s$ we write $E^{t:s}$ for the sequence $E^s$ minus its initial segment $E^t$.

Recall that an hypothesis and posterior scheme are given by a function $P_{(\cdot)}$ that for given sequence $E^t$ returns a distribution $P_{E^t} = P_{E^t}(\cdot \mid \Theta_{E^t})$ over hypothesis set $\Theta_{E^t}$.

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

This induces the distribution $P_{E^t}(\cdot) = \sum_{H \in \Theta_{E^t}} P_{E^t}(H) \cdot H(\cdot \mid E^t)$ over events in the outcome space.

The conditional distributions $P_{E^t}(\cdot \mid \Theta)$ for $\Theta \subseteq \Theta_{E^t}$ are clearly well-defined. One can also derive from the specifications of any of the open-minded versions we discussed that for $E^s \succ E^t$

$$P_{E^s}(\cdot \mid \Theta_{E^t}) = P_{E^t}(\cdot \mid E^{t:s}, \Theta_{E^t}), \tag{A.1}$$

a fact that we will rely on in the proofs of Lemma 1.15 and Corollary 1.16, in Section 6.5.

The conditional distributions $P_{E^t}(\cdot \mid \Theta)$ for $\Theta \supset \Theta_{E^t}$ are *not* well-defined, because the posteriors of the elements in $\Theta \setminus \Theta_{E^t}$ are not defined. Nevertheless, for the purpose of analyzing an open-minded agent's procedure of retro-actively setting a prior (as in the proof of Lemma 1.18 in Section 6.5), it will be useful to agree on the following. For $E^s \succ E^t$, the probability $P_{E^t}(H \mid \Theta_{E^s})$ is the posterior probability of $H \in \Theta_{E^s}$ after $E^t$, retroactively calculated from the posterior probability $P_{E^s}(H \mid \Theta_{E^s})$ after $E^s$. More precisely, we can define for all $H \in \Theta_{E^s}$,

$$P_{E^t}(H \mid E^{t:s}, \Theta_{E^s}) := P_{E^s}(H \mid \Theta_{E^s}), \tag{A.2}$$

from which the function $P_{E^t}(\cdot \mid \Theta_{E^s})$, by using the likelihoods of all $H \in \Theta_{E^s}$ on $E^{t:s}$, can unambiguously be retrieved.

***6.2. Calculations for Example 3.8.*** We want to ensure (15), that is,

$$\frac{P_0(H_{N_i} \mid \Theta_{N_i+1}) \cdot H_{N_i}(E^{t_i})}{\sum_{H \in \Theta_{N_i+1}} P_0(H \mid \Theta_{N_i+1}) \cdot H(E^{t_i})} > r. \tag{A.3}$$

Write $q := P_0(H_{N_i} \mid \Theta_{N_i+1})$ for the conditional prior, that by (13) equals

$$\frac{P_0(H_i)}{1 - \tau_{N_i+1}} = \frac{2^{-i} \cdot \tau_{N_0+1}}{1 - (1 - \sum_{j=1}^{i} 2^{-j}) \cdot \tau_{N_0+1}} = \frac{2^{-i} \cdot \tau_{N_0+1}}{1 - 2^{-i} \cdot \tau_{N_0+1}}. \tag{A.4}$$

Since $H_{N_i}(E^{t_i}) = 1$, (A.3) translates into

$$q > r \cdot \left( q + \sum_{H \in \Theta_{N_i+1} \setminus \{H_{N_i}\}} P_0(H \mid \Theta_{N_i+1}) \cdot H(E^{t_i}) \right), \tag{A.5}$$

that is,

$$\frac{1-r}{r} \cdot q > \sum_{H \in \Theta_{N_i+1} \setminus \{H_{N_i}\}} P_0(H \mid \Theta_{N_i+1}) \cdot H(E^{t_i}). \tag{A.6}$$

Now assuming that there is positive $\delta$ such that all other hypotheses' predictive probabilities are no more than $1 - \delta$ for each possible outcome from $t_{i-1}$ up to $t_i$, so that

$$\sum_{H \in \Theta_{N_i+1} \setminus \{H_{N_i}\}} P_0(H \mid \Theta_{N_i+1}) \cdot H(E^{t_i}) < (1-q) \cdot (1-\delta)^{t_i - t_{i-1}}, \tag{A.7}$$

it suffices for (A.6) that

$$\frac{1-r}{r} \cdot \frac{q}{1-q} > (1-\delta)^{t_i - t_{i-1}}. \tag{A.8}$$

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

Writing out

$$\frac{q}{1-q} = \frac{\left(\frac{2^{-i}\cdot\tau_{N_0+1}}{1-2^{-i}\cdot\tau_{N_0+1}}\right)}{\left(1-\frac{2^{-i}\cdot\tau_{N_0+1}}{1-2^{-i}\cdot\tau_{N_0+1}}\right)} = \frac{\left(\frac{2^{-i}\cdot\tau_{N_0+1}}{1-2^{-i}\cdot\tau_{N_0+1}}\right)}{\left(\frac{1}{1-2^{-i}\cdot\tau_{N_0+1}}\right)} = 2^{-i}\cdot\tau_{N_0+1}, \tag{A.9}$$

we thus require

$$\frac{1-r}{r}\cdot 2^{-i}\cdot\tau_{N_0+1} > (1-\delta)^{t_i-t_{i-1}}, \tag{A.10}$$

that is,

$$t_i - t_{i-1} > \frac{-\log(1-r) - (-\log r) + i - \log\tau_{N_0+1}}{-\log(1-\delta)}. \tag{A.11}$$

***6.3. Proof of Proposition 4.9.*** Let the truth $H^* \in \Theta$ be Bernoulli-1/2, and put $P(H^*) = 1/2$. Define an infinite series of times $t_0, t_1, t_2, \ldots$ by $t_0 = 0$, $t_{i+1} = t_i + i + 3$. For each time $t_i$, let $E^{t_i}_j$ be the $j$-th ($0 < j \leq 2^{t_i}$) outcome sequence of length $t_i$. We will now define a countable collection of hypotheses $H_{i,j}$ that each overfit one particular sequence between two successive times $t_{i-1}$ and $t_i$, and follow $H^*$ elsewhere. More precisely, we define for each $i$, for each positive $j \leq 2^{t_i}$ and the corresponding $j'$ such that $E^{t_{i-1}}_{j'} \prec E^{t_i}_j$, the hypothesis $H_{i,j}$ by

$$H_{i,j}(E^s) = \begin{cases} 2^{-t_{i-1}} & \text{if } E^{t_{i-1}}_{j'} \preccurlyeq E^s \preccurlyeq E^{t_i}_j \\ 0 & \text{if } E^{t_{i-1}}_{j'} \preccurlyeq E^s \text{ but } E^s \mid E^{t_i}_j \\ H^*(E^s)\cdot 2^{t_i - t_{i-1}} & \text{if } E^{t_i}_j \prec E^s \\ H^*(E^s) & \text{otherwise.} \end{cases} \tag{A.12}$$

Given an infinite outcome stream $E^\omega$. We can now assign positive prior to each of these hypotheses as follows. Denote by $(E^{t_i}_j)^{\mathrm{C}}$ the sequence $E^{t_i}_j$ with the very last outcome inverted, 0 for 1 or vice versa. For each $i$, for each $j \leq 2^{t_i}$, let

$$P(H_{i,j}) = \begin{cases} 2^{-i-2} & \text{if } (E^{t_i}_j)^{\mathrm{C}} \prec E^\omega \\ 2^{-i-2}\cdot(2^{t_i}-1)^{-1} & \text{otherwise.} \end{cases} \tag{A.13}$$

This is a valid prior assignment because $\sum_{H\in\Theta} P(H) = 2^{-1} + \sum_{i>0}(2^{-i-1}) = 1$.

Now we consider, on the stream $E^\omega$, for arbitrary $i$ and the $j$ such that $E^{t_i}_j \prec E^\omega$, the error in the agent's predictive probability $P(0 \mid E^{t_i-1}_j)$ after having observed all of $E^{t_i}_j$ but the very last outcome. That is, we consider the distance

$$\left| P(0 \mid E^{t_i-1}_j) - H^*(0 \mid E^{t_i-1}_j) \right|. \tag{A.14}$$

To this end, write $\Theta' := \Theta \setminus \{H_{i,j}\}$ and first consider the posterior ratio of $P(H_{i,j} \mid E^{t_i-1}_j)$, write $\alpha$, and $P(\Theta' \mid E^{t_i-1}_j) = 1-\alpha$,

$$\frac{\alpha}{1-\alpha} = \frac{P(H_{i,j} \mid E^{t_i-1}_j)}{P(\Theta' \mid E^{t_i-1}_j)} = \frac{P(H_{i,j})\cdot H_{i,j}(E^{t_i-1}_j)}{P(\Theta')\cdot P(E^{t_i-1}_j \mid \Theta')}. \tag{A.15}$$

It follows from specification (A.12) that all hypotheses in $\Theta'$ assign true probability $H^*(E^{t_i-1}_j)$ to $E^{t_i-1}_j$, except for the overfitting hypotheses $H_{i',j'}$ for $i' \leq i$ and $j'$ such that

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

there is $j''$ with $E_{j''}^{t_{i'}-1} \prec E_{j'}^{t_{i'}}, E^{\omega}$. But for each $i' < i$, among these hypotheses $H_{i',j'}$ there is only one $H_{i',k'}$ that does *not* give probability 0 to $E_j^{t_i-1}$, and with assignment (A.13) each member of the majority already holds at least as much prior as the single exception $H_{i',k'}$. Similarly, for $i$, it is, among these $H_{i,j'}$ and apart from $H_{i,j}$, only the hypothesis $H_{i,k}$ for $E_k^{t_i} \prec E^{\omega}$ that does not assign probability 0 to $E_j^{t_i-1}$, and each other $H_{i,j'}$ already holds at least as much prior as $H_{i,k}$. We thus have that the likelihood of hypothesis set $\Theta'$ satisfies

$$P(E_j^{t_i-1} \mid \Theta') = \sum_{H \in \Theta'} P(H \mid \Theta') \cdot H(E_j^{t_i-1}) < H^*(E_j^{t_i-1}) = 2^{-t_i+1}, \tag{A.16}$$

wherefore

$$\begin{aligned} \frac{\alpha}{1-\alpha} &> \frac{2^{-i-2} \cdot 2^{-t_{i-1}}}{(1-2^{-i-2}) \cdot 2^{-t_i+1}} \\ &= \frac{2^{-i-3}}{(1-2^{-i-2}) \cdot 2^{-(t_i-t_{i-1})}} \\ &= \frac{2^{-i-3}}{(1-2^{-i-2}) \cdot 2^{-i-3}} \\ &> 1, \end{aligned}$$

meaning that $\alpha > 1/2$.

Finally, apart from $H_{i,j}$, it is only the hypothesis $H_{i,k}$ for $E_k^{t_i} \prec E^{\omega}$ that is still included in the posterior over $\Theta$ conditional on $E_j^{t_i-1}$ (that did not assign probability 0 to $E_j^{t_i-1}$) *and* that gives a predictive probability $H_{i,k}(0 \mid E_j^{t_i-1})$ different from $H^*(0 \mid E_j^{t_i-1}) = 1/2$. Write $\alpha' := P(H_{i,k} \mid E_j^{t_i-1})$ for the posterior of $H_{i,k}$, and abbreviate $\Theta_{i;j,k} := \{H_{i,j}, H_{i,k}\}$. Since indeed $H_{i,k}(0 \mid E_j^{t_i-1}) = 1 - H_{i,j}(0 \mid E_j^{t_i-1})$,

$$P(0 \mid E_j^{t_i-1}, \Theta_{i;j,k}) = \frac{\alpha}{\alpha+\alpha'} \cdot H_{i,j}(0 \mid E_j^{t_i-1}) + \frac{\alpha'}{\alpha+\alpha'} \cdot H_{i,k}(0 \mid E_j^{t_i-1}) \tag{A.17}$$

evaluates to either $\frac{\alpha}{\alpha+\alpha'} = 1 - \frac{\alpha'}{\alpha+\alpha'}$ or $\frac{\alpha'}{\alpha+\alpha'}$. Using $\alpha' < 1/2 < \alpha$, it follows that

$$\left| P(0 \mid E_j^{t_i-1}, \Theta_{i;j,k}) - H^*(0 \mid E_j^{t_i-1}) \right| = 1/2 - \frac{\alpha'}{\alpha+\alpha'}. \tag{A.18}$$

We can then rewrite (A.14) as

$$\left| (\alpha+\alpha') \cdot P(0 \mid E_j^{t_i-1}, \Theta_{i;j,k}) + (1-(\alpha+\alpha')) \cdot H^*(0 \mid E_j^{t_i-1}) - H^*(0 \mid E_j^{t_i-1}) \right|, \tag{A.19}$$

which simplifies to

$$\begin{aligned} (\alpha+\alpha') \cdot \left| P(0 \mid E_j^{t_i-1}, \Theta_{i;j,k}) - H^*(0 \mid E_j^{t_i-1}) \right| &= (\alpha+\alpha') \cdot \left( 1/2 - \frac{\alpha'}{\alpha+\alpha'} \right) \\ &= \frac{\alpha+\alpha'}{2} - \alpha' \\ &> 1/4 - 1/2 \cdot \alpha'. \end{aligned}$$

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

But note that $H_{i,j}$ and $H_{i,k}$ have the same likelihood $H_{i,j}(E_j^{t_i-1}) = H_{i,k}(E_j^{t_i-1})$, so that by assignment (A.13) the ratio

$$\frac{\alpha}{\alpha'} = \frac{P(H_{i,j})}{P(H_{i,k})} = 2^{t_i} - 1, \tag{A.20}$$

which implies that $\alpha' < (2^{t_i} - 1)^{-1}$ is arbitrarily small for large enough $i$. That means that indeed for any choice of $\varepsilon > 0$, we have for infinitely many $i$ that

$$\left| P(0 \mid E_j^{t_i-1}) - H^*(0 \mid E_j^{t_i-1}) \right| > 1/4 - \varepsilon,$$

blocking convergence on the stream $E^\omega$.

***6.4. Proof of Proposition 4.10.*** Consider Example 3.8 with $t_0 = 0$, $\varepsilon' > 1/4$, and where after each $t_i$ all hypotheses $H_{N_j}$ for $j \leq i$ always give predictive probabilities $(1/2, 1/2)$. Let the sequence of time points $t_0 < t_1 < t_2 \ldots$ at which overfitting hypotheses are introduced satisfy (16), with prior assignments given by (13). This defines a hypothesis and posterior scheme, and thus induces a completed agent measure.

Next, take an infinite outcome stream $E^\omega$ that is constructed as follows. For any $i \geq 0$, take for the subsequence $E^{t_i+2:t_{i+1}}$ any sequence of length $t_{i+1} - t_i - 1$, and let $E_{t_i+1}$ be the outcome with $P_{t_i}(E_{t_i+1} \mid \Theta_{E^{t_i}}) < 1/2 - \varepsilon' = 1/4$ (for $E_1$ take either 0 or 1). Now the completed agent measure $P^\infty$ fails to hold a grain of $H^*$ on any such sequence $E^\omega$. Namely, for such a sequence $E^\omega$ we have by construction that for each $t$, with $j$ maximal such that $t_j < t$, that

$$P^\infty(E^t) < \left(2^{-1}\right)^{t-j} \cdot \left(2^{-2}\right)^j = 2^{-t-j}. \tag{A.21}$$

But since $2^{-t-j}/2^{-t} = 2^{-j}$ goes to 0 as $t$ hence $j$ goes to infinity, there is no positive $a$ such that $P^\infty(E^t) \geq a \cdot H^*(E^t)$ for all $t$.

***6.5. Proof of Theorems 4.11 and 4.13.*** We show for both the forward-looking open-minded Bayesian agent and its proto-version that for any hypothesis and posterior scheme, any finite outcome sequence $E^{t_0}$, for any hypothesis $H \in \Theta_{E^{t_0}}$, there is a constant $a \in (0, 1)$ such that for every outcome sequence $E^t \succcurlyeq E^{t_0}$ it holds that

$$P^\infty(E^{t_0:t} \mid E^{t_0}) \geq a \cdot H(E^{t_0:t} \mid E^{t_0}). \tag{A.22}$$

In words, for any outcome sequence $E^{t_0}$, the completed agent measure conditional on $E^{t_0}$ holds a positive grain of every hypothesis $H$ in $\Theta_{E^{t_0}}$. In particular, the completed agent measure conditional on $E^{t_0}$ holds a grain of the truth $H^*$, if $H^*$ is in $\Theta_{E^{t_0}}$.

Our proof consists of two main steps. First, we show that for any open-minded agent the completed agent measure conditional on $E^{t_0}$ dominates the agent function $P_{E^{t_0}}$ with a factor that involves the posteriors assigned to new hypotheses (Lemma 1.15 and Corollary 1.16). Second, we show for (the proto-version of) the forward-looking open-minded Bayesian that this latter factor is indeed at least a positive constant (Lemmas 1.17 and 1.18, respectively).

In all of the following statements we quantify over all $E^{t_0}$ and $E^t \succcurlyeq E^{t_0}$, and in the accompanying proofs we start by presupposing any such two sequences. This allows for the following simplified notation, that unambiguously pertains to a particular instantiated $E^t$ and initial segment $E^{t_0}$. We abbreviate $P_s := P_{E^s}$ and $\Theta_s := \Theta_{E^s}$ for all $E^s \preccurlyeq E^t$. Moreover, we always let $i \geq 0$ denote the number of new hypotheses that

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

are formulated along the sequence $E^{t_0+1:t}$, and we write $p_j := P_{t_j}(H_{E^{t_j}} \mid \Theta_{t_j})$ for the conditional posterior assigned to the $j$-th ( $j \leq i$) such hypothesis $H_{E^{t_j}} \in \Theta_{t_j} \setminus \Theta_{t_{j-1}}$, incoming at $t_j$.

LEMMA 1.15. *For an open-minded agent, we have that for any hypothesis and posterior scheme, for every $E^{t_0}$, every $E^t \succ E^{t_0}$, every $0 \leq j \leq i$,*

$$P_{t_j}(E^{t_j:t} \mid \Theta_{t_j}) \geq \frac{\prod_{k=0}^{j-1}(1-p_{k+1}) \cdot P_{t_0}(E^{t_0:t} \mid \Theta_{t_0})}{\prod_{k=0}^{j-1} P_{t_k}(E^{t_k:t_{k+1}} \mid \Theta_{t_k})}. \tag{A.23}$$

*Proof.* We proceed by induction. The base case, $j = 0$, follows trivially from empty products. Next, assuming as induction hypothesis that (A.23) holds for given $j < i$, we derive for $j+1$ that

$$\begin{aligned}
P_{t_{j+1}}(E^{t_{j+1}:t} \mid \Theta_{t_{j+1}}) &= \sum_{H \in \Theta_{t_{j+1}}} P_{t_{j+1}}(H \mid \Theta_{t_{j+1}}) \cdot H(E^{t_{j+1}:t} \mid E^{t_{j+1}}) \\
&\geq (1-p_{j+1}) \sum_{H \in \Theta_{t_j}} P_{t_j}(H \mid E^{t_j:t_{j+1}}, \Theta_{t_j}) \cdot H(E^{t_{j+1}:t} \mid E^{t_{j+1}}) \\
&= (1-p_{j+1}) \sum_{H \in \Theta_{t_j}} \frac{P_{t_j}(H \mid \Theta_{t_j}) \cdot H(E^{t_j:t_{j+1}} \mid E^{t_j})}{P_{t_j}(E^{t_j:t_{j+1}} \mid \Theta_{t_j})} \cdot H(E^{t_{j+1}:t} \mid E^{t_{j+1}}) \\
&= (1-p_{j+1}) \cdot \frac{\sum_{H \in \Theta_{t_j}} P_{t_j}(H \mid \Theta_{t_j}) \cdot H(E^{t_j:t} \mid E^{t_j})}{P_{t_j}(E^{t_j:t_{j+1}} \mid \Theta_{t_j})} \\
&= \frac{(1-p_{j+1}) \cdot P_{t_j}(E^{t_j:t} \mid \Theta_{t_j})}{P_{t_j}(E^{t_j:t_{j+1}} \mid \Theta_{t_j})} \\
&\geq \frac{(1-p_{j+1}) \cdot \prod_{k=0}^{j-1}(1-p_{k+1}) \cdot P_{t_0}(E^{t_0:t} \mid \Theta_{t_0})}{P_{t_j}(E^{t_j:t_{j+1}} \mid \Theta_{t_j}) \cdot \prod_{k=0}^{j-1} P_{t_k}(E^{t_k:t_{k+1}} \mid \Theta_{t_k})} \\
&= \frac{\prod_{k=0}^{j}(1-p_{k+1}) \cdot P_{t_0}(E^{t_0:t} \mid \Theta_{t_0})}{\prod_{k=0}^{j} P_{t_k}(E^{t_k:t_{k+1}} \mid \Theta_{t_k})}.
\end{aligned}$$

□

COROLLARY 1.16. *For an open-minded agent, we have that for any hypothesis and posterior scheme, for every $E^{t_0}$, every $E^t \succ E^{t_0}$,*

$$P^{\infty}(E^{t_0:t} \mid E^{t_0}) \geq \prod_{j=0}^{i-1}(1-p_{j+1}) \cdot P_{t_0}(E^{t_0:t} \mid \Theta_{t_0}). \tag{A.24}$$

*Proof.* We write out

$$\begin{aligned}
P^{\infty}(E^{t_0:t} \mid E^{t_0}) &= \prod_{s=t_0}^{t-1} P_s(E_{s+1} \mid \Theta_s) \\
&= \left( \prod_{j=0}^{i-1} \prod_{s=t_j}^{t_{j+1}-1} P_s(E_{s+1} \mid \Theta_s) \right) \prod_{s=t_i}^{t-1} P_s(E_{s+1} \mid \Theta_{t_i})
\end{aligned}$$

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

$$= \left( \prod_{j=0}^{i-1} P_{t_j}(E^{t_j:t_{j+1}} \mid \Theta_{t_j}) \right) \cdot P_{t_i}(E^{t_i:t} \mid \Theta_{t_i}),$$

where the latter equality follows from the fact that for each $j$ and $t_j \leq t'_j < t_{j+1}$ we have

$$\begin{aligned} \prod_{s=t_j}^{t'_j} P_s(E_{s+1} \mid \Theta_s) &= \prod_{s=t_j}^{t'_j} P_{t_j}(E_{s+1} \mid E^{t_j:s}, \Theta_{t_j}) \\ &= \prod_{s=t_j}^{t'_j} \frac{P_{t_j}(E^{t_j:s+1} \mid \Theta_{t_j})}{P_{t_j}(E^{t_j:s} \mid \Theta_{t_j})} \\ &= \frac{P_{t_j}(E^{t_j:t_{j+1}} \mid \Theta_{t_j})}{P_{t_j}(E^{t_j:t_j} \mid \Theta_{t_j})} \\ &= P_{t_j}(E^{t_j:t_{j+1}} \mid \Theta_{t_j}). \end{aligned}$$

But applying Lemma (A.23) for $i = j$ then yields

$$\begin{aligned} P^{\infty}(E^t \mid E^{t_0}) &\geq \left( \prod_{j=0}^{i-1} P_{t_j}(E^{t_j:t_{j+1}} \mid \Theta_{t_j}) \right) \cdot \frac{\prod_{j=0}^{i-1}(1 - p_{j+1}) \cdot P_{t_0}(E^t \mid \Theta_{t_0})}{\prod_{j=0}^{i-1} P_{t_j}(E^{t_{j+1}} \mid \Theta_{t_j})} \\ &= \prod_{j=0}^{i-1}(1 - p_{j+1}) \cdot P_{t_0}(E^t \mid \Theta_{t_0}). \end{aligned}$$

□

LEMMA 1.17. *For the proto-version of the forward-looking open-minded agent, we have that for every hypothesis and posterior scheme, there is a constant $b \in (0, 1)$ such that for every $E^{t_0}$, every $E^t \succ E^{t_0}$,*

$$\prod_{j=1}^{i}(1 - p_j) \geq b. \tag{A.25}$$

*Proof.* We have by specification that $0 < p_j \leq c$ for each $j$ and a positive constant $c < 1$, and that $\sum_{j=1}^{i} p_j \leq d$ for some positive constant $d$. Using the standard inequality $\frac{x-1}{x} \leq \ln x$ for $x > 0$, this allows us to derive

$$\begin{aligned} -\ln \prod_{j=1}^{i}(1 - p_j) &= \sum_{j=1}^{i} -\ln(1 - p_j) \\ &\leq \sum_{j=1}^{i} \frac{p_j}{1 - p_j} \\ &\leq \frac{1}{1 - c} \sum_{j=1}^{i} p_j \\ &\leq \frac{d}{1 - c}, \end{aligned}$$

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press

where the second inequality follows from the fact that $1 - c \leq 1 - p_j$ for all $j$. Thus we have

$$\prod_{j=1}^{i}(1 - p_j) \geq \exp\left(-\frac{d}{1-c}\right), \tag{A.26}$$

yielding the desired statement with constant $b = \exp\left(-\frac{d}{1-c}\right)$ independent of $E^t$. □

LEMMA 1.18. *For the forward-looking open-minded agent, we have that for every hypothesis and posterior scheme, there is a constant $b \in (0,1)$ such that for every $E^{t_0}$, every $E^t \succ E^{t_0}$,*

$$\prod_{j=1}^{i}(1 - p_j) \geq b. \tag{A.27}$$

*Proof.* By specification, and in particular the abstention trick (26), for each $j$ the posterior $p_j = P_{t_j}(H_{t_j} \mid \Theta_{t_j})$ conditional on $\Theta_{t_j}$ equals the prior $P_0(H_{t_j} \mid \Theta_{t_j})$ conditional on $\Theta_{t_j}$. But the latter is calculated from a choice of *absolute* prior, denoted $p'_j$, by

$$p_j = \frac{p'_j}{1 - \tau_j} = \frac{\tau_{j-1} - \tau_j}{1 - \tau_j}, \tag{A.28}$$

where $\tau_j$ is the probability of the catch-all after formulation of $H_{t_j}$. We thus have

$$\begin{aligned}\prod_{j=1}^{i}(1 - p_j) &= \prod_{j=1}^{i}\left(1 - \frac{\tau_{j-1} - \tau_j}{1 - \tau_j}\right)\\ &= \prod_{j=1}^{i}\left(\frac{1 - \tau_{j-1}}{1 - \tau_j}\right)\\ &= \frac{1 - \tau_0}{1 - \tau_i}\\ &\geq 1 - \tau_0,\end{aligned}$$

yielding the desired statement with constant $b = 1 - \tau_0$ independent of $E^t$. □

Finally, combining the previous results, we obtain that for the (proto-version of) the forward-looking open-minded Bayesian, for any hypothesis and posterior scheme, any $E^{t_0}$, any hypothesis $H \in \Theta_{E^{t_0}}$, any $E^t \succcurlyeq E^{t_0}$, it holds that

$$\begin{aligned}P^{\infty}(E^{t_0:t} \mid E^{t_0}) &\geq \prod_{j=0}^{i-1}(1 - p_{j+1}) \cdot P_{t_0}(E^{t_0:t} \mid \Theta_{t_0})\\ &\geq b \cdot P_{t_0}(E^{t_0:t} \mid \Theta_{t_0})\\ &\geq b \cdot P_{t_0}(H \mid \Theta_{t_0}) \cdot H(E^{t_0:t} \mid E^{t_0}),\end{aligned}$$

yielding the desired statement (A.22) with constant $a = b \cdot P_{t_0}(H \mid \Theta_{t_0})$ independent of $E^{t_0:t}$. □

**Acknowledgments.** We thank Wouter Koolen for many valuable discussions and suggestions, Jan-Willem Romeijn and Sylvia Wenmackers for their feedback on a first version of this paper, and Jan Sprenger for his commentary on an early version of this

paper at the Formal Epistemology Workshop in Turin. Further thanks go to audiences in Amsterdam, Geneva, Munich, and Turin, and to Jean Baccelli, Kathleen Creel, Sam Fletcher, Konstantin Genin, Peter Grünwald, Stephan Hartmann, Simon Huttegger, Jonathan Livengood, Daniel Malinsky, Conor Mayo-Wilson, Miklós Rédei, Rush Stewart, Kino Zhao, and the anonymous reviewers.

## BIBLIOGRAPHY

[1] Barnard, G. A. & Cox, D. R., editors. (1962). *The Foundations of Statistical Inference*: *A Discussion*. Methuen's Monographs on Applied Probability and Statistics. London: Methuen.

[2] Belot, G. (2013). Bayesian orgulity. *Philosophy of Science*, **80**(4), 483–503.

[3] Blackwell, D. & Dubins, L. (1962). Merging of opinion with increasing information. *The Annals of Mathematical Statistics*, **33**, 882–886.

[4] Cesa-Bianchi, N. & Lugosi, G. (2006). *Prediction, Learning and Games*. Cambridge: Cambridge University Press.

[5] Chernov, A. & Vovk, V. G. (2009). Prediction with expert evaluators' advice. In Gavaldà, R., Lugosi, G., Zeugmann, T., and Zilles, S., editors. *Proceedings of the 20th International Conference on Algorithmic Learning Theory*. Lecture Notes in Computer Science, Vol. 5809. Berlin: Springer, pp. 8–22.

[6] Chihara, C. S. (1987). Some problems for Bayesian confirmation theory. *British Journal for the Philosophy of Science*, **38**(4), 551–560.

[7] Dawid, A. P. (1982). The well-calibrated Bayesian. *Journal of the American Statistical Association*, **77**(379), 605–610.

[8] ———. (1985). The impossibility of inductive inference. Comment on Oakes (1985). *Journal of the American Statistical Association*, **80**(390), 340–341.

[9] Earman, J. (1992). *Bayes or Bust*? *A Critical Examination of Bayesian Confirmation Theory*. A Bradford Book. Cambridge, MA: MIT Press.

[10] Freund, Y., Schapire, R. E., Singer, Y., & Warmuth, M. K. (1997). Using and combining predictors that specialize. In *Proceedings of the 29th Annual ACM Symposion on Theory of Computing*. New York: ACM Press, pp. 334–343.

[11] Gaifman, H. & Snir, M. (1982). Probabilities over rich languages, testing and randomness. *Journal of Symbolic Logic*, **47**(3), 495–548.

[12] Gelman, A. & Shalizi, C. R. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, **66**, 8–38.

[13] Gillies, D. A. (2001). Bayesianism and the fixity of the theoretical framework. In Corfield, D. and Williamson, J., editors, *Foundations of Bayesianism*. Applied Logic Series, Vol. 24. Dordrecht: Springer, pp. 363–379.

[14] Glymour, C. (2016). Responses to the contributions to the special issue "Causation, probability, and truth: The philosophy of Clark Glymour". *Synthese*, **193**(4), 1251–1285.

[15] Howson, C. (1988). On the consistency of Jeffreys's simplicity postulate, and its role in Bayesian inference. *The Philosophical Quarterly*, **38**(150), 68–83.

[16] ———. (2000). *Hume's Problem*: *Induction and the Justification of Belief*. New York: Oxford University Press.

[17] Huttegger, S. M. (2015). Merging of opinions and probability kinematics. *Review of Symbolic Logic*, **8**(4), 611–648.

[18] Jeffreys, H. (1961). *Theory of Probability* (third edition). Oxford: Oxford University Press.

[19] Kalai, E. & Lehrer, E. (1993). Rational learning leads to Nash equilibrium. *Econometrica* **61**(5), 1019–1045.

[20] ———. (1994). Weak and strong merging of opinions. *Journal of Mathematical Economics*, **23**(1), 73–86.

[21] Koolen, W. M., Adamskiy, D. & Warmuth, M. K. (2012). Putting Bayes to sleep. In Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., editors. *Proceedings of the 25th International Conference on Neural Information Processing Systems*. Red Hook, NY: Curran Associates Inc., pp. 135–143.

[22] Lehrer, E. & Smorodinsky, R. (1996). Merging and learning. In Ferguson, T. S., Shapley, L. S., and MacQueen, J. B., editors. *Statistics, Probability and Game Theory*: *Papers in Honor of David Blackwell*. Lecture Notes—Monograph Series, Vol. 30. Hayward, CA: Institute of Mathematical Statistics, pp. 147–168.

[23] Leike, J. (2016). Nonparametric General Reinforcement Learning. PhD Dissertation, Australian National University.

[24] Lindley, D. V. (1982). Comment on Dawid (1982). *Journal of the American Statistical Association*, **77**(379), 611–612.

[25] ———. (2000). The philosophy of statistics. *Journal of the Royal Statistical Society*: *Series D* (*The Statistician*), **49**(3), 293–337.

[26] Mourtada, J. & Maillard, O.-A. (2017). Efficient tracking of a growing number of experts. In Hanneke, S. and Reyzin, L., editors. *Proceedings of the 28th International Conference on Algorithmic Learning Theory*. Proceedings of Machine Learning Research, Vol. 76. PMLR, pp. 517–539.

[27] Oakes, D. (1985). Self-calibrating priors do not exist. *Journal of the American Statistical Association*, **80**(390), 339.

[28] Putnam, H. (1963). 'Degree of confirmation' and inductive logic. In Schilpp, P. A., editor. *The Philosophy of Rudolf Carnap*. The Library of Living Philosophers, Vol. XI. LaSalle, IL: Open Court, pp. 761–783.

[29] Raidl, E. (2020). Open-minded orthodox Bayesianism by epsilon-conditionalization. *British Journal for the Philosophy of Science*, **71**(1), 139–176.

[30] Romeijn, J.-W. (2004). Hypotheses and inductive predictions. *Synthese*, **141**(3), 333–364.

[31] ———. (2005). Theory change and Bayesian statistical inference. *Philosophy of Science*, **72**(5), 1174–1186.

[32] Shimony, A. (1969). Letter to Rudolf Carnap, September 20. Rudolf Carnap Papers, 1905–1970, ASP.1974.01, Series XII. Notes and Correspondence: Probability, Mathematics, Publishers, UCLA Administrative, and Lecture Notes, 1927–1970, Subseries 1: Probability Authors, Box 84b, Folder 55. Pittsburgh, PA: Special Collections Department, University of Pittsburgh.

[33] ———. (1970). Scientific inference. In Colodny, R. G., editor. *The Nature and Function of Scientific Theories*: *Essays in Contemporary Science and Philosophy*. University of Pittsburgh Series in the Philosophy of Science, Vol. 4. Pittsburgh, PA: University of Pittsburgh Press, pp. 79–172.

[34] Sprenger, J. (2020). Conditional degree of belief and Bayesian inference. *Philosophy of Science*, **87**(2), 319–335.

[35] Sprenger, J. & Hartmann, S. (2019). *Bayesian Philosophy of Science*. Oxford: Oxford University Press.

[36] Sterkenburg, T. F. (2019). Putnam's diagonal argument and the impossibility of a universal learning machine. *Erkenntnis*, **84**(3), 633–656.

[37] Vassend, O. B. (2019). New semantics for Bayesian inference: The interpretative problem and its solutions. *Philosophy of Science*, **86**(4), 696–718.

[38] Wenmackers, S. & Romeijn, J.-W. (2016). New theory about old evidence: A framework for open-minded Bayesianism. *Synthese*, **193**(4), 1225–1250.

MUNICH CENTER FOR MATHEMATICAL PHILOSOPHY
LUDWIG-MAXIMILIANS-UNIVERSITY MUNICH
GESCHWISTER-SCHOLL-PLATZ 1, 80539 MUNICH, GERMANY
*E-mail*: tom.sterkenburg@lmu.de

MACHINE LEARNING GROUP
CENTRUM WISKUNDE & INFORMATICA
SCIENCE PARK 123, 1098XG, THE NETHERLANDS
and
MATHEMATICAL INSTITUTE
LEIDEN UNIVERSITY
NIELS BOHRWEG 1, 2333CA LEIDEN, THE NETHERLANDS
*E-mail*: r.de.heide@cwi.nl

https://doi.org/10.1017/S1755020321000022 Published online by Cambridge University Press