



ARTICLE

Jeffrey Conditionalization and Value Changes

Ilho Park 

Department of Philosophy & Institute of Critical Thinking and Writing, Jeonbuk National University,
Jeonju-si, South Korea
Email: ipark.phil@gmail.com

(Received 13 December 2024; revised 17 March 2025; accepted 13 May 2025)

Abstract

Dmitri Gallow has recently proposed an ingenious accuracy-first vindication that our credence should be updated in accordance with conditionalization. This paper extends his idea to cases where we undergo only a partial learning experience. In particular, I attempt to vindicate, in the spirit of accuracy-first epistemology, that when we undergo what is called ‘Jeffrey partial learning’ in this paper, our credences should be updated in accordance with Jeffrey conditionalization, or at the very least, the update should be rigid. In doing so, I propose what I call the ‘Jeffrey-accuracy function.’ This function is not strictly proper and, at first glance, seems to rationalize ill-motivated credence updating. However, this turns out not to be the case.

Keywords: Jeffrey conditionalization; accuracy; value changes; Jeffrey-accuracy function

1. Conditionalization and value changes

For the past few decades, accuracy-firsters, who regard accuracy as the only ultimate epistemic value, have tried to vindicate various Bayesian epistemic norms such as probabilism, conditionalization, the Reflection Principle, and the Principal Principle. Some seem successful; some seem not. Their approach to probabilism is generally well accepted as an outstanding epistemic vindication of why our credences should be probabilistically coherent. However, the accuracy-first approach to conditionalization had seemed to run counter to orthodox Bayesianism – that is, this approach had appeared to lead to somewhat unanticipated results, such as denying conditionalization or appealing to an ill-founded decision-theoretic rule.¹

¹There is extensive literature containing accuracy-first projects that aim to vindicate various Bayesian norms. Among them, Pettigrew (2016) is one of the most representative. Some readers may disagree with my assessment of the accuracy-first approach to conditionalization. This may be, in particular, because certain works, such as Greaves and Wallace (2006), are widely recognized as successfully vindicating a version of conditionalization. However, readers should note that the discussion that follows is largely unrelated to their vindication of conditionalization. Generally, Greaves and Wallace (2006) is regarded as a vindication of a synchronic version of conditionalization, often referred to as ‘Plan Conditionalization.’ See Easwaran (2013) and Pettigrew (2016). In this paper, I will focus not on the synchronic version of conditionalization but on its diachronic version. Of course, there have been several attempts to vindicate this version, for example,

Fortunately, Gallow (2019) recently provided an ingenious accuracy-first way to get out of this predicament. His breakthrough idea is that what is changed as a result of learning experiences is the way credences are epistemically valued – not credences *per se*. According to his suggestion, the changes in credences are rationalized by the changes in epistemic values and the decision-theoretic principle of maximizing expected epistemic values.

To see this, let me introduce several notations and terminology. Let ' \mathcal{V} ' be an epistemic value function. This function takes a credence function c and a world w as input and returns a real number representing the epistemic value of c at w . Relatedly, accuracy-firsters take it that the epistemic value of c at w , i.e., $\mathcal{V}(c, w)$, is entirely determined by its proximity to the truth function of w – in other words, the epistemic value of c at w is identified with its accuracy at that world. In what follows, I will use interchangeably 'epistemic values' and 'accuracies'.

With this accuracy function \mathcal{V} in hand, the accuracy-firsters provide a decision-theoretic way of rationalizing our credence updating. Let ' $E_p(c, \mathcal{V})$ ' be the expected accuracy of c by the light of p .² This is defined as follows:

$$E_p(c, \mathcal{V}) := \sum_{w \in \mathcal{W}} p(w) \mathcal{V}(c, w).$$

This is a weighted average of the accuracies of c at each world w , with a weight being your credence that w is actual. Now, an accuracy-first updating rule, which may be called 'the principle of maximization of the expected accuracy (MEA)', can be formulated, as follows:

MEA. Suppose that an accuracy function \mathcal{V} is legitimate, and that an agent's credence function p is probabilistically coherent. Then, the agent's credence updating from p to p' is rational if it holds that:

$$p' = \arg \max_c [E_p(c, \mathcal{V})]. \quad (1)$$

It is noteworthy that MEA might rationalize ill-motivated credence updating unless the accuracy function \mathcal{V} is constrained in a reasonable way. Suppose, for example, that there is an accuracy function \mathcal{V} such that $E_p(p, \mathcal{V}) \leq E_p(p^*, \mathcal{V})$ but $p \neq p^*$. According to MEA, then, an agent, whose credence function is p , could update their credence function to p^* . Notably, this updating is still rationalized by MEA even though no new relevant evidence is obtained.

Accuracy-firsters, of course, do not need to worry about this point. Several constraints have been imposed on legitimate accuracy functions. Among these constraints, Strict Propriety can play a role in preventing the ill-motivated updating.³

Strict Propriety. Suppose that an accuracy function \mathcal{V} is legitimate, and that a credence function p is probabilistically coherent. Then, for any credence function c ($\neq p$),

Leitgeb and Pettigrew (2010) and Pettigrew (2016). As Gallow (2019) points out, the vindication in Leitgeb and Pettigrew (2010) relies on an ill-founded decision-theoretic rule – in particular, the rule in question cannot, mathematically speaking, be regarded as a weighted average. On the other hand, the attempt in Pettigrew (2016) leads to a diachronic rule that differs from conditionalization – namely, what is called 'Brute Laplacian Imaging.'

²Every credence function in this paper is assumed to be defined over the same opinion set, which is also assumed to be the power set of a finite, non-empty set of possible worlds \mathcal{W} . Thus, propositions are regarded as members of the power set. For the sake of notational simplicity, I often use w rather than $\{w\}$ if there is no danger of confusion. For example, I will use $p(w)$ instead of $p(\{w\})$.

³There are several arguments for Strict Propriety. Some discussions about the relationship between Strict Propriety and ill-motivated credence updating can be found in Gibbard (2007), Joyce (2009), and Oddie (1997).

$$E_p(p, \mathcal{V}) > E_p(c, \mathcal{V}).$$

This constraint states that a legitimate accuracy function leads us to expect our own credence function to be valued more highly than any other credence functions. Thus, if \mathcal{V} is strictly proper, MEA does not rationalize that a credence function is updated to another function despite the absence of relevant evidence. With the help of Strict Propriety, our credence function can rationally remain unchanged until we obtain some new evidence.

What about when we do obtain some evidence? Orthodox Bayesians require us to update our credence function in accordance with conditionalization:

Conditionalization. Suppose that an agent learns that E is true, and nothing more, thereby updating their credence function from a probabilistically coherent credence function p to another function p_E . Then, p_E results from conditionalizing p on E if and only if $p_E(-) = p(-|E)$, where $p(E) > 0$.

Can accuracy-firsters vindicate, by appealing to MEA, the rationality of credence updating in accordance with conditionalization? If \mathcal{V} is strictly proper, then any credence updating, except for trivial cases, cannot be rationalized by MEA, and therefore, conditionalization cannot be vindicated either.⁴ Thus, accuracy-firsters need to find ways to modify MEA to make it a rule governing our rational credence updating.

In this regard, Gallow's proposal can essentially be understood as abandoning the idea that we should always evaluate credence functions using strictly proper accuracy functions. Gallow thinks, especially, that when we learn something, a learning-encoded accuracy function, even if it is not strictly proper, should be used to evaluate credence functions. According to him, accuracy-firsters, who aim to vindicate any diachronic rule in terms of the single-minded pursuit of accuracy, should prioritize changes in value rather than changes in credence – in other words, accuracy-firsters should take it that learning experience rationalizes changes of accuracy functions and such changes, in turn, rationalize changes in credence functions.

Then, what function should play the role of a learning-encoded function after E is learned? Suppose that an agent who evaluates credence functions using a strictly proper accuracy function \mathcal{V} learns that E is true and nothing more. Gallow proposes that, in this situation, the accuracy function used to evaluate credence functions should be changed to \mathcal{V}_E , defined as follows:

$$\mathcal{V}_E(c, w) := \begin{cases} \mathcal{V}(c, w) & \text{if } w \in E \\ k & \text{if } w \in \neg E \end{cases}$$

Here, k is a constant. According to this proposal, when E is learned, some epistemic values remain the same while others do not. In particular, the epistemic values at the worlds where E is false change so that they attain the same value as each other.

What intuition does this kind of change in epistemic value capture? To see this, consider the following story about practical value changes due to a learning experience. You have a date with your partner this weekend. You were told that they would book one of the following for this date: a jazz concert, a rock concert, a baseball game, or a basketball game. So, you began to care about these possibilities and assigned a practical value to each. Sometime later, you learn that your partner has not booked any music concert for this weekend. Thus, the possibility of having a date at a music concert is eliminated, so you no longer take its value into account for your weekend date. In this

⁴Here, 'trivial cases' denote situations where the relevant agent learns something that they already know or learns a tautology, thereby leaving their credence function unchanged.

way, the learning experience can change the extent to which the values of some possibilities are taken into account, and this change, in turn, affects your practical value of the possibilities. In particular, if such possibilities have a practical value, then their values should be the same. This is because when the practical values of the possibilities are not taken into account, any differences in their practical value are canceled out. Basically, Gallow's proposal captures the same intuition as the one reflected in the story above. When E is learned, the possible worlds where E is false are eliminated. As a result, the epistemic values of a credence function at those worlds are not taken into account, which leads to a change in the epistemic values so that they become the same.

This learning-encoded accuracy function \mathcal{V}_E leads accuracy-firsters to a new version of MEA, which can be applied to cases where we obtain some evidence.

MEA1. Suppose that a learning-encoded accuracy function \mathcal{V}_E is legitimate. Suppose also that an agent learns that E is true, and nothing more, thereby updating their credence function from a probabilistically coherent credence function p to another function p_E . Then, the agent's credence updating from p to p_E is rational if it holds that:

$$p_E = \arg \max_c [E_p(c, \mathcal{V}_E)]. \quad (2)$$

It is not hard to prove that when \mathcal{V}_E is generated by a strictly proper function \mathcal{V} , equation (2) holds if and only if p_E results from conditionalizing p on E – that is, $p_E(-) = p(-|E)$.⁵ Thus, conditionalization can be said, in the spirit of accuracy-first epistemology, to be a rationality requirement governing our credence updating.

Before we proceed further, one remark about Gallow's proposal is in order. As mentioned, the learning-encoded accuracy function \mathcal{V}_E is not strictly proper. Indeed, it is the case that:

$$E_p(p, \mathcal{V}_E) < E_p(p(-|E), \mathcal{V}_E),$$

according to which an agent whose credence function is p expects their own credence function to be valued less than $p(-|E)$ when epistemic values are given by \mathcal{V}_E . Does Gallow's proposal rationalize the aforementioned ill-motivated credence updating? Gallow (2019, 19–20) does not think so, and I agree.

One way to see this may be to note that: for any credence function $c (\neq p_E)$,⁶

$$E_{p_E}(p_E, \mathcal{V}_E) > E_{p_E}(c, \mathcal{V}_E),$$

where \mathcal{V} is an accuracy function generated by a strictly proper accuracy function \mathcal{V} , and p_E results from conditionalizing p on E . This indicates that \mathcal{V}_E leads an agent to expect their own posterior function p_E to be valued more highly than any other credence functions. Hence, \mathcal{V}_E can be said to be strictly proper in a restricted sense – that is, \mathcal{V}_E is strictly proper with respect to credence functions that result from conditionalization on E . So, if our attention is restricted to such credence functions, it seems natural to say that \mathcal{V}_E , which is generated by a strictly proper accuracy function \mathcal{V} , does not lead to ill-motivated credence updating, and thus is at least as legitimate as strictly proper \mathcal{V} . A caveat needs to be stated here. I think this type of response will, in Section 4, turn out not to be sufficiently general to account for reasons that will become clear later. I will revisit this issue in Section 3.⁷

⁵A relevant proof can be found in Gallow (2019), 18–19.

⁶It is assumed that \mathcal{V} is strictly proper and p_E is probabilistically coherent. Note also that $p_E(w) = 0$ when $w \notin E$. Then, we have that for any credence function $c (\neq p_E)$, $\sum_{w \in \mathcal{W}} p_E(w) \mathcal{V}_E(p_E, w) = \sum_{w \in \mathcal{W}} p_E(w) \mathcal{V}(p_E, w) > \sum_{w \in \mathcal{W}} p_E(w) \mathcal{V}(c, w) = \sum_{w \in \mathcal{W}} p_E(w) \mathcal{V}_E(c, w)$.

⁷In particular, this type of response will, in Section 4, turn out not to be sufficiently general to account for cases where we undergo a partial learning experience. In that section, I will argue that the partial learning-

2. Jeffrey partial learning and value changes

Heretofore, we have considered doxastic situations in which experience leads us to learn a single proposition, and our credences are updated accordingly. However, it may not be the case that experience always leads us to learn a single proposition. Due to our conceptual and/or cognitive limitations, experience may often fall short of leading to such learning. Be that as it may, such experience, which I will henceforth refer to as ‘partial learning’, cannot be said to have no impact on our doxastic states. Partial learning may change some aspect of our doxastic state, and its impact, in turn, may propagate throughout our overall state.

In this regard, it is noteworthy that some orthodox Bayesians have considered a particular type of partial learning, which may be called ‘Jeffrey partial learning’. This expression is intended to denote a course of partial learning experience that changes our doxastic state over a partition, which is a set of mutually exclusive and collectively exhaustive propositions, rather than a single proposition. For orthodox Bayesians, such experience directly shifts our credence assignment over a partition, and its impact leads to an update of our overall credences in accordance with Jeffrey conditionalization.

Let $\mathbb{E} = \{E_1, \dots, E_m\}$ be a partition. In what follows, I will say that an agent undergoes Jeffrey partial learning on \mathbb{E} with the input parameters $\beta_{i,1}$, exactly when the Jeffrey partial learning shifts their credence in E_i from e_i to e_i^+ for any E_i in \mathbb{E} . Here, $\beta_{i,1} = (e_i^+/e_i)/(e_1^+/e_1)$, called the ‘Bayes factor of E_i against E_1 ’ (where E_1 is an arbitrary anchored proposition).⁸ The term ‘input parameters’ is intended to denote parameters representing the impact of experience on credences with prior credences factored out. Many orthodox Bayesians consider Bayes factors to be among the most plausible input parameters.⁹ With these input parameters in hand, Jeffrey conditionalization can be formulated as follows:

Jeffrey Conditionalization. Suppose that an agent undergoes Jeffrey partial learning on \mathbb{E} with the input parameters $\beta_{i,1}$, and nothing more, thereby updating their credence function from a probabilistically coherent credence function p to another function $p_{\mathbb{E}}$. Then, $p_{\mathbb{E}}$ results from Jeffrey conditionalizing p on \mathbb{E} with $\beta_{i,1}$ if and only if

$$p_{\mathbb{E}} = \frac{\sum_i \beta_{i,1} p(-\&E_i)}{\sum_i \beta_{i,1} p(E_i)}. \quad (3)$$

As is widely known, equation (3) holds only if the credence updating from p to $p_{\mathbb{E}}$ is rigid with respect to \mathbb{E} in the following sense:

Rigidity.

$$p_{\mathbb{E}}(-|E_i) = p(-|E_i), \text{ for any } E_i \in \mathbb{E}.$$

That is to say, credence updating in accordance with Jeffrey conditionalization ensures that the conditional credences given each proposition in \mathbb{E} remain unchanged. Rigidity is regarded as an essential part of Jeffrey conditionalization.

Can accuracy-firsters then vindicate that an agent should update their credence function in accordance with Jeffrey conditionalization when they undergo Jeffrey partial

encoded accuracy function is not strictly proper even in this restricted sense, and propose an epistemological way to address the issue related to Strict Propriety and ill-motivated updating. See footnote 14.

⁸It is assumed here that e_i and e_i^+ are all positive real numbers in $(0, 1)$.

⁹For the discussion about Bayes factors and the input parameters, see Field (1978), Jeffrey (2004), and Wagner (2002); Wagner (2003). In those works, the input parameters are often called ‘observational parameters’, ‘probabilistic observational reports’, ‘indices of probability change’, and so on.

learning \mathbb{E} ? Specifically, can they provide a Gallow-style vindication of Jeffrey conditionalization in terms of the single-minded pursuit of accuracy? Accuracy-firsters, tasked with providing such a vindication, should maintain that when an agent undergoes Jeffrey partial learning on \mathbb{E} , they should evaluate credence functions using an accuracy function that encodes the partial learning, rather than the accuracy function that the agent used before. In what follows, I will refer to such an accuracy function as ‘Jeffrey-accuracy function on \mathbb{E} .’

Then, what should the Jeffrey-accuracy function on \mathbb{E} look like? A simple but Gallow-style suggestion might be that when an agent, who evaluates credence functions using a strictly proper accuracy function \mathcal{V} , undergoes Jeffrey partial learning on \mathbb{E} and nothing more, the Jeffrey-accuracy function $\mathcal{V}_{\mathbb{E}}$, defined as follows, should be used to evaluate credence functions: for any $E_i \in \mathbb{E}$,

$$\mathcal{V}_{\mathbb{E}}(c, w) := \lambda_i \mathcal{V}(c, w) + k_i \quad \text{if } w \in E_i.$$

Here, each λ_i is a positive real number and each k_i is a real number.

The Jeffrey-accuracy function $\mathcal{V}_{\mathbb{E}}$ captures a similar intuition to Gallow’s learning-encoded accuracy function \mathcal{V}_E . Consider a story that is identical with the date story in the previous section, except that you learn your partner has not booked any music concert for this weekend. Instead, you undergo an experience that leads you to think that your partner may strongly prefer attending a sports game over a music concert this weekend. Unlike the original story, this experience does not eliminate any possibilities. However, the experience can also have an impact on the way the possibilities are practically valued. After undergoing the experience, you may pay more attention to things related to attending a sports game than to those related to attending a music concert – for example, you may come to take more time to choose an outfit suitable for a sports game than for a music concert. This shows that you take the value of attending a sports game into account much more than that of attending a music concert, and that, as a result, the practical value of each changes.

In this way, partial learning changes the way possibilities are practically valued. Similar changes occurs in epistemic value. A Jeffrey partial learning can eliminate no possible world. However, such learning changes the extent to which the epistemic value of a credence function at each world is taken into account, and thus the epistemic value of the function at each world. Each λ_i in the definition of $\mathcal{V}_{\mathbb{E}}(c, w)$ can be taken as a weight representing the extent to which an agent takes the epistemic value of c at the world w where E_i is true into account. Before undergoing a Jeffrey partial learning, the agent assigns the same weight to the epistemic value of c at each world. However, these weights are adjusted when the agent undergoes partial learning. In what follows, each weight λ_i will be called ‘the accuracy factor of E_i .’ On the other hand, each k_i in the definition can be taken as a constant epistemic value, which is assigned to c at the world w where E_i is true, even when E_i turns out to be false and λ_i becomes zero.

Similar to the learning-encoded accuracy function \mathcal{V}_E , the Jeffrey-accuracy function on \mathbb{E} , i.e., $\mathcal{V}_{\mathbb{E}}$, leads accuracy-firsters to another version of MEA, which can be applied to cases where we undergo Jeffrey partial learning on a partition \mathbb{E} :¹⁰

¹⁰In orthodox decision problems, utilities are not affected by experience. Experience has an impact on the expected utilities only through credences reflecting the experience. In MEA2, however, experience directly affects the way credences are epistemically valued—that is, experience directly alters the accuracy function. Thus, in MEA2, experience has an impact on the expected value in two different ways: first, through the prior credence function p , reflecting the old experience, and second, through the new accuracy function $\mathcal{V}_{\mathbb{E}}$, reflecting the new experience. Similar considerations apply to MEA1. In this sense, MEA2 and MEA1 can be thought of as unorthodox decision rules.

MEA2. Suppose that a Jeffrey-accuracy function $\mathcal{V}_{\mathbb{E}}$ is legitimate. Suppose also that an agent undergoes Jeffrey partial learning on \mathbb{E} , and nothing more, thereby updating their credence function from a probabilistically coherent credence function p to another function $p_{\mathbb{E}}$. Then, the credence updating from p to $p_{\mathbb{E}}$ is rational if it holds that:

$$p_{\mathbb{E}} = \arg \max_c [E_p(c, \mathcal{V}_{\mathbb{E}})]. \quad (4)$$

Interestingly, we can prove that equation (4) holds if and only if¹¹

$$p_{\mathbb{E}} = \frac{\sum_i \lambda_i p(-\&E_i)}{\sum_i \lambda_i p(E_i)}, \quad (5)$$

where $\mathcal{V}_{\mathbb{E}}$ is generated by a strictly proper function \mathcal{V} .

What conclusion can accuracy-firsters, who aim to vindicate Jeffrey conditionalization in terms of the single-minded pursuit of accuracy, draw from this result? Pay attention to the striking similarity between the two equations, (3) and (5) – in particular, note that by setting the Bayes factor $\beta_{i,1}$ in equation (3) to the ratio of the two accuracy factors, λ_i and λ_1 , i.e., λ_i/λ_1 , we arrive at equation (5). Thus, such accuracy-firsters might think as follows: Jeffrey partial learning on \mathbb{E} rationalizes the change from \mathcal{V} to $\mathcal{V}_{\mathbb{E}}$, and such changes, in turn, rationalize our credence updating in accordance with Jeffrey conditionalization on \mathbb{E} , using the ratios of the accuracy factors λ_i/λ_1 as input parameters.

Admittedly, this kind of vindication might be unsatisfactory to orthodox Bayesians. Such Bayesians might argue that experience directly changes our credences themselves, rather than the way credences are epistemically valued. Thus, they may require accuracy-firsters to explain how it is rationalized that the ratio of the accuracy factors in question serves as input parameters that represent the impact of experience on credences with prior credences factored out.

Is there any accuracy-first way to respond to this requirement? In particular, how can accuracy-firsters, who aim to vindicate Jeffrey conditionalization in terms of the single-minded pursuit of accuracy, respond to this requirement? I think they have only a few things in this regard. Similar to what is stated in Gallow (2019, 17), it could be argued that there is no compelling reason to favor the explanation provided by orthodox Bayesians about the impact of experience over that of the single-minded accuracy-firsters. However, I think it must be acknowledged that this seems to fall short of being a full-fledged response capable of persuading orthodox Bayesians.

Nonetheless, this problem does not render the aforementioned accuracy-first argument for Jeffrey conditionalization entirely useless. Note that the credence updating from p to $p_{\mathbb{E}}$ satisfies equation (5) only if the credence updating from p to $p_{\mathbb{E}}$ is rigid with respect to \mathbb{E} . Thus, it can be said that the accuracy-first approach to Jeffrey partial learning successfully vindicates that our credence updating should be rigid when we undergo Jeffrey partial learning and nothing more. As stated, Rigidity is an essential part of Jeffrey conditionalization. Therefore, the aforementioned accuracy-first vindication can be seen as providing a good epistemic reason why it is rational for us to update our credences in accordance with Jeffrey conditionalization.

¹¹Proofs and mathematical remarks related to the following discussion are given in Appendix.

3. Jeffrey-accuracy and ill-motivated updating

So, we can say that accuracy-firsters, at the very least, vindicate that our credence updating should be rigid when we undergo Jeffrey partial learning. That said, this vindication remains incomplete. Can Jeffrey-accuracy function $\mathcal{V}_{\mathbb{E}}$ be regarded as a legitimate epistemic value function? Is $\mathcal{V}_{\mathbb{E}}$ strictly proper? As mentioned in Section 1, it can be argued that the learning-encoded value function \mathcal{V}_E , which is generated by a strictly proper function \mathcal{V} , is strictly proper in a restricted sense. Does a similar argument apply to $\mathcal{V}_{\mathbb{E}}$? To put it another way, can we say that $\mathcal{V}_{\mathbb{E}}$ is strictly proper with respect to credence functions that result from Jeffrey conditionalization on \mathbb{E} ? Unfortunately, it does not.

Suppose that an accuracy function \mathcal{V} is strictly proper and that $\mathcal{V}_{\mathbb{E}}$ is generated by \mathcal{V} with the accuracy factors λ_i and the constants k_i . Let p be a probabilistically coherent function and let $p^{(n)}$ be a function defined as follows:

$$p^{(n)} := \frac{\sum_i (\lambda_i)^n p(-\&E_i)}{\sum_i (\lambda_i)^n p(E_i)}.$$

Notably, it can be proved that:¹² for any $n \in \{0, 1, \dots\}$,

$$p^{(n+1)} = \arg \max_c \left[E_{p^{(n)}}(c, \mathcal{V}_{\mathbb{E}}) \right].$$

Let $p_{\mathbb{E}}$ be a credence function resulting from Jeffrey conditionalizing p on \mathbb{E} with the input parameters λ_i/λ_1 . So, it holds that $p_{\mathbb{E}} = p^{(1)}$. Then, it follows from the above equation that:

$$E_{p_{\mathbb{E}}}(p_{\mathbb{E}}, \mathcal{V}_{\mathbb{E}}) < E_{p_{\mathbb{E}}}(p^{(2)}, \mathcal{V}_{\mathbb{E}}). \quad (6)$$

Hence, we should say that $\mathcal{V}_{\mathbb{E}}$ is not strictly proper. Moreover, unlike \mathcal{V}_E , the Jeffrey-accuracy function $\mathcal{V}_{\mathbb{E}}$ is not strictly proper, even in a restricted sense – that is, $\mathcal{V}_{\mathbb{E}}$ is not strictly proper even with respect to credence functions, such as $p_{\mathbb{E}}$, which result from Jeffrey conditionalizing p on \mathbb{E} .

Unfortunately, this is not the whole story regarding the Jeffrey-accuracy function $\mathcal{V}_{\mathbb{E}}$. The above result indicates that $\mathcal{V}_{\mathbb{E}}$ leads an agent to expect their own credence function $p_{\mathbb{E}}$ to be valued less than another credence function, namely $p^{(2)}$. Thus, an agent who follows MEA2 should update their credence function to $p^{(2)}$, even if there is no extra experience prompting this update. In other words, MEA2 rationalizes ill-motivated credence updates, provided that $\mathcal{V}_{\mathbb{E}}$ is legitimate. Note further that: for any $n \in \{0, 1, \dots\}$,

$$E_{p^{(n)}}(p^{(n)}, \mathcal{V}_{\mathbb{E}}) < E_{p^{(n)}}(p^{(n+1)}, \mathcal{V}_{\mathbb{E}}).$$

Hence, the ill-motivated credence updating cannot help but be iterated infinitely, and so $\mathcal{V}_{\mathbb{E}}$ renders our doxastic state very unstable. Based on the above considerations, we also reach a somewhat perplexing conclusion. Let E^+ be the disjunction of all propositions whose accuracy factor has the maximum value. Then, we have that $p^{(n)}$ will converge to $p(-|E^+)$ as n increases.¹³ Put differently, as the ill-motivated credence updating in question is iterated,

¹²Proofs and mathematical remarks related to the following discussion are given in Appendix.

¹³Formally, E^+ is defined as follows: $E^+ = \vee_i \{E_i : \lambda_i \geq \lambda_j, \text{ for any } j\}$. Let λ^+ be the maximum value of accuracy factors. And let $I^+ = \{i : \lambda_i \geq \lambda_j, \text{ for any } j\}$. Then, we have that:

$$\lim_{n \rightarrow \infty} p^{(n)} = \lim_{n \rightarrow \infty} \frac{\sum_i (\lambda_i)^n p(-\&E_i)}{\sum_i (\lambda_i)^n p(E_i)} = \lim_{n \rightarrow \infty} \frac{\sum_i (\lambda_i/\lambda^+)^n p(-\&E_i)}{\sum_i (\lambda_i/\lambda^+)^n p(E_i)}$$

the agent's credence function will approach $p(-|E^+)$. When Jeffrey partial learning on \mathbb{E} legitimately changes epistemic value functions as suggested, the credence updating guided by MEA2 collapses into conditionalization on E^+ !

What conclusion should we draw from the above considerations? Should we abandon the accuracy-first vindication of Jeffrey conditionalization? Or, should we conclude that MEA2 is theoretically redundant and seek a way to live our epistemic lives without it? It seems to me that these options are not on the right track. I think, in particular, that the above problems are merely apparent and not genuine.

Suppose that you are at the beginning of your epistemic life, so you have not yet undergone any (partial) learning experience. Let a probability function p be your initial credence function reflecting this doxastic situation. How should you epistemically evaluate credence functions in this situation? In particular, is it reasonable for you to evaluate the epistemic betterness between credence functions using a Jeffrey-accuracy function $\mathcal{V}_{\mathbb{E}}$ in this situation? It is very natural that *your evaluation of credence functions should reflect your doxastic situation exactly as it stands*. If you evaluate credence functions using $\mathcal{V}_{\mathbb{E}}$ encoding experience that you do not undergo, you might reach a wrong verdict about epistemic betterness, thereby updating your credence function even in the absence of relevant experience. So, at the beginning of the epistemic life, you should evaluate credence functions using an accuracy function \mathcal{V} that does not encode any experience.

Similar things go with a doxastic situation where you undergo Jeffrey partial learning on \mathbb{E} , and nothing more. When you undergo such an experience, the accuracy function you should use for epistemic evaluation is the Jeffrey-accuracy function $\mathcal{V}_{\mathbb{E}}$, which encodes your experience exactly as it stands. Neither \mathcal{V} , which encodes no experience, nor \mathcal{V}_E , which encodes a learning experience you did not undergo, should be used for your epistemic evaluation. In this doxastic situation, the epistemic evaluation of a credence function c is thus made by considering the expected Jeffrey-accuracy of c , which is the weighted average of the Jeffrey-accuracy of c at each world w , with the weight being your credence that w is actual. Note that the evaluation is conducted from the perspective of you, who have not yet updated the credence function despite undergoing the experience in question. Thus, the weight assigned to the Jeffrey-accuracy of c at each world w must be given by your initial function p , which does not reflect the impact of Jeffrey partial learning on your credences. To wrap up, your evaluation in this situation depends on two things: the Jeffrey-accuracy function $\mathcal{V}_{\mathbb{E}}$ and your initial credence function p . The impact of the experience on your doxastic state is reflected only in the former, not in the latter.

Suppose now that you have completed your credence updating after undergoing the Jeffrey partial learning in question, and thus have $p_{\mathbb{E}}$ as your *current* credence function. With $p_{\mathbb{E}}$ as your current credence function, you evaluate credence functions by considering their expected accuracies by the light of $p_{\mathbb{E}}$. In this evaluation, the weight assigned to the accuracy of a credence function at a world w is your *current* credence that w is actual, i.e., $p_{\mathbb{E}}(w)$. Thus, it should be said that when you currently evaluate credence functions by the light of $p_{\mathbb{E}}$, the weights reflect the impact of Jeffrey partial learning, which prompted your previous credence update to $p_{\mathbb{E}}$, on your doxastic state.

Then, what accuracy function should be used for this current epistemic evaluation? Should you make use of the Jeffrey-accuracy function $\mathcal{V}_{\mathbb{E}}$, which encodes your previous Jeffrey partial learning experience? I do not think so. As explained, the weights that are

$$= \frac{\sum_{i \in I^+} p(-\&E_i)}{\sum_{i \in I^+} p(E_i)} = \frac{p(-\&E^+)}{p(E^+)} = p(-|E^+).$$

needed to determine the expected accuracy of credence functions – i.e., your current credence given by p_E —already reflect the impact of the experience. Thus, if you currently evaluate a credence function c by considering its expected Jeffrey-accuracy in light of p_E , i.e., $E_{p_E}(c, \mathcal{V}_E)$, your verdict based on this evaluation should be said to be *overdetermined*. This is because the evaluation considers the impact of the experience on your doxastic state twice: once through p_E and once through \mathcal{V}_E . Our evaluation of credence functions should reflect our doxastic situation exactly as it stands. Any evaluation that leads us to such an overdetermined verdict reflects our doxastic situation incorrectly and should therefore be regarded as unreasonable. To avoid this kind of unreasonableness, you should, in this doxastic situation, use an accuracy function \mathcal{V} that encodes no impact of experience, rather than the Jeffrey-accuracy function \mathcal{V}_E . To sum up, when p_E is your current credence function, your current epistemic evaluation of a credence function c should be made by considering $E_{p_E}(c, \mathcal{V})$, not $E_{p_E}(c, \mathcal{V}_E)$.

Let me revisit the problems mentioned above: the Jeffrey-accuracy function \mathcal{V}_E is not strictly proper and rationalizes ill-motivated credence updating. I concede the former; however, I do not concede the latter. As shown in (6), it is an indisputable mathematical fact that \mathcal{V}_E is not strictly proper. Nonetheless, this fact does not directly lead us to the conclusion that \mathcal{V}_E rationalizes ill-motivated credence updating. This is because, as explained above, \mathcal{V}_E is not a legitimate accuracy function for an agent to use when they have a credence function p_E that already reflects the impact of Jeffrey partial learning. The Jeffrey-accuracy function \mathcal{V}_E is legitimate only when we have not yet updated our credence function. Hence, the epistemic evaluation using $E_{p_E}(c, \mathcal{V}_E)$ cannot be reasonable, and therefore, such an evaluation does not rationalize ill-motivated credence updating. We do not need to worry about whether the Jeffrey-accuracy function makes the credence updating guided by MEA2 collapse into conditionalization.¹⁴

4. Conclusion

Gallow (2019) has proposed that our learning experience rationalizes changes in the way credences are epistemically valued, and such changes, in turn, rationalize our credence updating in accordance with conditionalization. This paper may be considered an extension of his idea to cases where our experience falls short of leading to such learning. In particular, I have attempted to vindicate, in the spirit of accuracy-first epistemology, that when we undergo Jeffrey partial learning, our credences should be updated in accordance with Jeffrey conditionalization, or at the very least, the update should be rigid.

In doing so, I have formulated Jeffrey-accuracy functions, which are generated by strictly proper accuracy functions. The strictly proper function that an agent has before undergoing Jeffrey partial learning assigns the same weight to the epistemic value of a credence function at each world. These weights are adjusted when the agent undergoes

¹⁴Can this line of response also apply to the issue related to a learning-encoded accuracy function \mathcal{V}_E and Strict Propriety, as discussed toward the end of Section 2? Of course, it can. I do concede that \mathcal{V}_E is not strictly proper—this is a mathematical fact. Nonetheless, I do not concede that \mathcal{V}_E rationalizes ill-motivated credence updating. Suppose p_E is your current credence function, obtained by conditionalizing p on E , and \mathcal{V}_E is generated by a strictly proper accuracy function \mathcal{V} . In this doxastic situation, your current epistemic evaluation of a credence function c should be based on $E_{p_E}(c, \mathcal{V})$, not $E_{p_E}(c, \mathcal{V}_E)$. Thus, no ill-motivated credence updating is rationalized by \mathcal{V}_E . Since your current credence function p_E already encodes the impact of the learning experience, \mathcal{V}_E is not legitimate for use in the epistemic evaluation of credence functions in such a doxastic situation. Here, it is noteworthy that $E_{p_E}(c, \mathcal{V}) = E_{p_E}(c, \mathcal{V}_E)$. Using this mathematical fact, we could assert toward the end of Section 2 that \mathcal{V}_E is strictly proper in a restricted sense. However, as stated in footnote 7, I think this line of response is not promising. Instead, it is more epistemologically plausible to focus on which accuracy function should be used given a particular doxastic situation, as explained in this section.

Jeffrey partial learning. I have called such weights ‘accuracy factors’. These factors represent the extent to which the agent takes the accuracy of a credence function at each world into account. As discussed, Jeffrey-accuracy functions are not strictly proper and, at first glance, seem to rationalize ill-motivated credence updating. Fortunately, this turns out not to be the case. Therefore, this paper can be regarded as providing an accuracy-first vindication of Jeffrey conditionalization and/or rigidity.

I would like to finish this paper by mentioning the striking similarity between the Bayes factors and the accuracy factors. According to orthodox Bayesians, the Bayes factors are regarded as representing the impact of experience with prior credences factored out. Notably, the discussion of accuracy factors in this paper seems to provide a new way of interpreting the Bayes factors—namely, an accuracy-first interpretation of Bayes factors. As discussed, the Bayes factor of E_i against E_1 in Jeffrey conditionalization is the same as the ratio of the accuracy factor of E_i to that of E_1 in MEA2. Thus, the Bayes factors can be interpreted as representing the relative extent to which the accuracy of a credence function at each world is taken into account. This interpretation of Bayes factors may be good news for accuracy-firsters who aim to vindicate various epistemic norms in terms of the single-minded pursuit of accuracy.

Acknowledgments. I should be grateful to Namjoong Kim, Jaemin Jung, and Minkyung Wang for their valuable comments. I should also thank the anonymous reviewers of this journal for their suggestions and comments. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2023S1A5A2A01081804).

References

- Easwaran K. (2013). ‘Expected Accuracy Supports Conditionalization—and Conglomerability and Reflection.’ *Philosophy of Science* 80(1), 119–42.
- Field H. (1978). ‘A Note on Jeffrey Conditionalization.’ *Philosophy of Science* 45(3), 361–67.
- Gallow J.D. (2019). ‘Learning and Value Change.’ *Philosophers’ Imprint* 19, 1–22.
- Gibbard A. (2007). ‘Rational Credence and the Value of Truth.’ In T.S. Gendler and J. Hawthorne (eds), *Oxford Studies in Epistemology: Volume 2: Volume 2*. Oxford: Oxford University Press.
- Greaves H. and Wallace D. (2006). ‘Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility.’ *Mind* 115(459), 607–32.
- Jeffrey R. (2004). *Subjective Probability: The Real Thing*. Cambridge: Cambridge University Press.
- Joyce J. (2009). ‘Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief.’ In F. Huber and C. Schmidt-Petri (eds), *Degrees of Belief*, pp. 263–97. London: Synthese.
- Leitgeb H. and Pettigrew R. (2010). ‘An Objective Justification of Bayesianism ii: The Consequences of Minimizing Inaccuracy.’ *Philosophy of Science* 77(2), 236–72.
- Oddie G. (1997). ‘Conditionalization, Cogency, and Cognitive Value.’ *British Journal for the Philosophy of Science* 48(4), 533–41.
- Pettigrew R. (2016). *Accuracy and the Laws of Credence*. Oxford: Oxford University Press UK.
- Wagner C.G. (2002). ‘Probability Kinematics and Commutativity.’ *Philosophy of Science* 69(2), 266–78.
- Wagner C.G. (2003). ‘Commuting Probability Revisions: The Uniformity Rule.’ *Erkenntnis* 59(3), 349–64.

Appendix

In this appendix, I will prove several mathematical results that the discussion in this paper depends on. Let p be a probabilistically coherent credence function, and let $\mathbb{E} = \{E_1, \dots, E_m\}$ be a partition. In what follows, I will assume that $p(E_i) > 0$ for any $E_i \in \mathbb{E}$. Now, let each λ_i be a positive real number. A credence function $p^{(n)}$ is defined as follows:

$$p^{(n)} := \frac{\sum (\lambda_i)^n p(-\&E_i)}{\sum_i (\lambda_i)^n p(E_i)},$$

where $n \in \{0, 1, \dots\}$. Then, we can show the following lemma.

LEMMA. For any $n \in \{0, 1, \dots\}$, $p^{(n)}$ is probabilistically coherent.

Proof. Note that $p^{(0)} = p$. So, $p^{(0)}$ is probabilistically coherent. For our purpose, it is sufficient to show that if $p^{(n)}$ is probabilistically coherent, then so is $p^{(n+1)}$. Note that it follows from the definition of $p^{(n)}$ that:

$$\begin{aligned} p^{(n+1)} &= \frac{\sum_j (\lambda_j)^{n+1} p(-\&E_j)}{\sum_j (\lambda_j)^{n+1} p(E_j)} = \frac{\sum_j \lambda_j (\lambda_j)^n p(-\&E_j)}{\sum_j \lambda_j (\lambda_j)^n p(E_j)} \\ &= \frac{\sum_j \lambda_j \frac{(\lambda_j)^n p(-\&E_j)}{\sum_i (\lambda_i)^n p(E_i)}}{\sum_j \lambda_j \frac{(\lambda_j)^n p(E_j)}{\sum_i (\lambda_i)^n p(E_i)}} = \frac{\sum_j \lambda_j \frac{\sum_i (\lambda_i)^n p(-\&E_j \& E_i)}{\sum_i (\lambda_i)^n p(E_i)}}{\sum_j \lambda_j \frac{\sum_i (\lambda_i)^n p(E_j \& E_i)}{\sum_i (\lambda_i)^n p(E_i)}} \\ &= \frac{\sum_j \lambda_j p^{(n)}(-\&E_j)}{\sum_j \lambda_j p^{(n)}(E_j)}. \end{aligned} \quad (A1)$$

Suppose now that $p^{(n)}$ is probabilistically coherent. It is trivial that $p^{(n+1)}(T) = 1$ and $p^{(n+1)}(X) \geq 0$, where T is a tautology and X is an arbitrary proposition. Consider two arbitrary propositions, X and Y , that are mutually exclusive. It follows from (A1) that:

$$\begin{aligned} p^{(n+1)}(X \vee Y) &= \frac{\sum_i (\lambda_i)^{n+1} p((X \vee Y) \& E_i)}{\sum_i (\lambda_i)^{n+1} p(E_i)} \\ &= \frac{\sum_i \lambda_i p^{(n)}((X \vee Y) \& E_i)}{\sum_i \lambda_i p^{(n)}(E_i)} \\ &= \frac{\sum_i \lambda_i p^{(n)}(X \& E_i)}{\sum_i \lambda_i p^{(n)}(E_i)} + \frac{\sum_i \lambda_i p^{(n)}(Y \& E_i)}{\sum_i \lambda_i p^{(n)}(E_i)} \\ &= p^{(n+1)}(X) + p^{(n+1)}(Y). \end{aligned}$$

Hence, we have that $p^{(n+1)}$ is probabilistically coherent. As a result, it can be concluded that for any $n \in \{0, 1, \dots\}$, $p^{(n)}$ is probabilistically coherent, as required.

With this lemma in hand, we can prove the following theorem, which contains what is mentioned in Section 3. The lemma and the following theorem are sufficient to demonstrate our claims given in the main text.

THEOREM. Suppose that \mathcal{V} is a strictly proper accuracy function, and that $\mathcal{V}_{\mathbb{E}}$ is generated by \mathcal{V} with λ_i and k_i , as follows: for any $E_i \in \mathbb{E}$,

$$\mathcal{V}_{\mathbb{E}}(c, w) = \lambda_i \mathcal{V}(c, w) + k_i, \text{ for any } w \in E_i.$$

Then, it holds that:

$$\operatorname{argmax}_c [E_{p^{(n)}}(c, \mathcal{V}_{\mathbb{E}})] = p^{(n+1)},$$

where $E_{p^{(n)}}(c, \mathcal{V}_{\mathbb{E}}) = \sum_{w \in \mathcal{W}} p^{(n)}(w) \mathcal{V}_{\mathbb{E}}(c, w)$.

Proof. Suppose that \mathcal{V} is a strictly proper accuracy function, and that $\mathcal{V}_{\mathbb{E}}$ is generated by \mathcal{V} with λ_i and k_i . Then, we obtain from (A1) that:

$$\begin{aligned} E_{p^{(n)}}(c, \mathcal{V}_{\mathbb{E}}) &= \sum_{w \in \mathcal{W}} p^{(n)}(w) \mathcal{V}_{\mathbb{E}}(c, w) = \sum_{E_i \in \mathbb{E}} \sum_{w \in E_i} p^{(n)}(w) (\lambda_i \mathcal{V}(c, w) + k_i) \\ &= \sum_{E_i \in \mathbb{E}} \sum_{w \in E_i} \lambda_i p^{(n)}(w) \mathcal{V}(c, w) + \sum_{E_i \in \mathbb{E}} \sum_{w \in E_i} k_i p^{(n)}(w) \\ &= \sum_{E_i \in \mathbb{E}} \sum_{w \in \mathcal{W}} \lambda_i p^{(n)}(w \& E_i) \mathcal{V}(c, w) + \sum_{E_i \in \mathbb{E}} k_i p^{(n)}(E_i) \end{aligned}$$

$$\begin{aligned}
&= \sum_{w \in \mathcal{W}} \sum_{E_i \in \mathbb{E}} \lambda_i p^{(n)}(w \& E_i) \mathcal{V}(c, w) + \sum_{E_i \in \mathbb{E}} k_i p^{(n)}(E_i) \\
&= \left(\sum_{E_i \in \mathbb{E}} \lambda_i p^{(n)}(E_i) \right) \sum_{w \in \mathcal{W}} \frac{\sum_{E_i \in \mathbb{E}} \lambda_i p^{(n)}(w \& E_i)}{\sum_{E_i \in \mathbb{E}} \lambda_i p^{(n)}(E_i)} \mathcal{V}(c, w) + \sum_{E_i \in \mathbb{E}} k_i p^{(n)}(E_i) \\
&= \left(\sum_{E_i \in \mathbb{E}} \lambda_i p^{(n)}(E_i) \right) \sum_{w \in \mathcal{W}} p^{(n+1)}(w) \mathcal{V}(c, w) + \sum_{E_i \in \mathbb{E}} k_i p^{(n)}(E_i) \\
&= \left(\sum_{E_i \in \mathbb{E}} \lambda_i p^{(n)}(E_i) \right) E_{p^{(n+1)}}(c, \mathcal{V}) + \sum_{E_i \in \mathbb{E}} k_i p^{(n)}(E_i).
\end{aligned}$$

Note that \mathcal{V} is assumed to be strictly proper, and that according to Lemma, $p^{(n+1)}$ is probabilistically coherent. Thus, we have that:

$$\begin{aligned}
\operatorname{argmax}_c [E_{p^{(n)}}(c, \mathcal{V}_{\mathbb{E}})] &= \operatorname{argmax}_c [E_{p^{(n+1)}}(c, \mathcal{V})] \\
&= p^{(n+1)},
\end{aligned}$$

as required.

Remark 1. Suppose that an agent whose credence function is p undergoes a Jeffrey partial experience on \mathbb{E} , and thus evaluates credence functions using $\mathcal{V}_{\mathbb{E}}$, which is generated by a strictly proper function \mathcal{V} . Then, the theorem and MEA2 imply that $p_{\mathbb{E}}$ should be equal to $p^{(1)}$, as explained in the main text. Note, in particular, that the theorem guarantees that the two equations (4) and (5) in the main text are equivalent to each other.

Remark 2. According to Lemma, $p^{(n)}$ and $p^{(n+1)}$ are both probabilistically coherent. Note that it is assumed that $p(E_i) > 0$ for any $E_i \in \mathbb{E}$. It is not hard to obtain from Lemma and this assumption that for any $n \in \{0, 1, \dots\}$, $p^{(n)}(-|E_i)$ s are all well-defined. On the other hand, (A1) implies that: for any $E_i \in \mathbb{E}$,

$$\begin{aligned}
p^{(n+1)}(E_i) &= \frac{\sum_j \lambda_j p^{(n)}(E_i \& E_j)}{\sum_j \lambda_j p^{(n)}(E_j)} = \frac{\lambda_i p^{(n)}(E_i)}{\sum_j \lambda_j p^{(n)}(E_j)}; \\
p^{(n+1)}(-\& E_i) &= \frac{\sum_j \lambda_j p^{(n)}(-\& E_i \& E_j)}{\sum_j \lambda_j p^{(n)}(E_j)} = \frac{\lambda_i p^{(n)}(-\& E_i)}{\sum_j \lambda_j p^{(n)}(E_j)}.
\end{aligned}$$

From this, thus, it follows that for any $n \in \{0, 1, \dots\}$, $p^{(n)}(-|E_i) = p^{(n+1)}(-|E_i)$ holds for any $E_i \in \mathbb{E}$ —in other words, for any $n \in \{0, 1, \dots\}$, the credence updating from $p^{(n)}$ to $p^{(n+1)}$ is rigid with respect to \mathbb{E} . This result, together with Theorem, jointly implies that *the credence updating from p to $p_{\mathbb{E}}$, which is guided by MEA2, is rigid with respect to \mathbb{E}* , as noted Section 2.

Remark 3. \mathcal{V} is assumed to be strictly proper, and it holds, according to the above proof, that $\operatorname{argmax}_c [E_{p^{(n)}}(c, \mathcal{V}_{\mathbb{E}})] = \operatorname{argmax}_c [E_{p^{(n+1)}}(c, \mathcal{V})] = p^{(n+1)}$. Thus, $p^{(n+1)}$ is the *unique* function that $E_{p^{(n+1)}}(c, \mathcal{V})$, and so $E_{p^{(n)}}(c, \mathcal{V}_{\mathbb{E}})$ render to have the maximum value. Therefore, it holds that $E_{p^{(n)}}(p^{(n)}, \mathcal{V}_{\mathbb{E}}) < E_{p^{(n)}}(p^{(n+1)}, \mathcal{V}_{\mathbb{E}})$, as explained Section 3.

Ilho Park is a professor in the Department of Philosophy and Institute of Critical Thinking and Writing at Jeonbuk National University in South Korea. His research primarily focuses on Bayesian epistemology and the philosophy of science. He has published several papers on accuracy-based epistemology, confirmation, objective chance, and related topics. Recently, he has been expanding his interests to the epistemology of machine learning—particularly, the interpretation of probability in deep neural network models.