Justification and update

Peijnenburg, Jeanne; Atkinson, David

Published in:
Asian Journal of Philosophy

DOI:
10.1007/s44204-024-00193-4

Publication date:
2024

Asian Journal of Philosophy (2024) 3:63
https://doi.org/10.1007/s44204-024-00193-4

BOOK SYMPOSIUM

Justification and update

Comments on *Thinking about Statistics* by Jun Otsuka

Jeanne Peijnenburg[1] · **David Atkinson**[1]

Received: 6 June 2024 / Accepted: 5 September 2024

Abstract
In this commentary on Jun Otsuka's first-rate book, we focus on the difference between justification and update.

1 Introduction

Thinking about Statistics by Jun Otsuka is a fine book, engagingly written and full of interesting details. Its subtitle, *The Philosophical Foundations*, might give the idea that we are dealing with philosophy of statistics, but the author makes clear that this would be a mistake: his aim is not to cover the "wealth of discussions concerning the theoretical ground of inductive inference, interpretations of probability, the everlasting battle between Bayesian and frequentist statistics, and so forth" (3). Nor is the book meant as an introduction, be it to statistics or to philosophy (ibid.), even though it contains lucid expositions on *p*-values, confidence levels, and significance tests, as well as instructive explanations of philosophical positions concerning probabilistic inference.

Rather, the book aims to be a bridge between debates in statistics and debates in philosophy, notably epistemology. Otsuka connects Bayesian statistics with epistemological internalism (Chapter 2), classical statistics with externalism (Chapter 3), model selection with pragmatism, and deep learning with virtue epistemology (both in Chapter 4). The way in which Otsuka builds this composite bridge is original and stimulating, but he also warns that the comparison should be handled with care:

✉ Jeanne Peijnenburg
jeanne.peijnenburg@rug.nl

David Atkinson
d.atkinson@rug.nl

1 Faculty of Philosophy, University of Groningen, Oude Boteringestraat 52, Groningen 9712 GL, Netherlands

Published online: 01 October 2024

> "By no means is this comparison intended to suggest that there is a perfect analogy between the statistical methodologies on the one hand and the philosophical doctrine on the other. The simple dichotomy proposed in this book hardly does justice to the nuanced and complicated practice of statistics and epistemology in the literature." (179)

Wise words, and in this short note, we will zoom in on what we think is an important difference between Bayesianism and internalist epistemology. This is not to deny, however, that parallels exist and can in fact be very fruitful.

2 Epistemology as the heart of statistics

Otsuka characterizes (inferential) statistics as "the art of inferring and estimating unobserved phenomena" (17). It starts with observed phenomena, and presupposes, in line with Hume's Uniformity of Nature, that patterns between observed phenomena also exist among unobserved ones. However, often in observed phenomena, disparate patterns can be detected, and we do not expect all of these patterns to continue in the future. Here, we encounter the infamous difference between natural laws and accidental generalizations, i.e., between statements that are and those that are not "lawlike," to use Nelson Goodman's term. Goodman tried to distinguish the two by arguing that only the predicates in lawlike statements are "entrenched," but this idea has been disputed.

In his book, Otsuka does not mention Goodman, yet the way in which he distinguishes between a "probability model" and a "statistical model" is reminiscent of the difference between Hume's point and that of Goodman. While a probability model embodies the uniformity of nature in that it goes beyond what is observed and serves as a theoretical ground for making inferences about unobserved data (41), statistical models offer testable specifications of this presumed uniformity. Otsuka calls them "probabilistic kinds." Examples are various probability distributions, such as the Bernoulli, the Poisson, or the normal distribution. Other examples are regression models, even though they do not represent patterns among observable data, but rather correlations between explanatory and response variables (113). And since a deep neural network is seen as "a gigantic regression model" (126), such a network, too, is an instance of a probabilistic kind—which indicates that Otsuka markedly expands Goodman's thoughts.

At the heart of statistics lies epistemology (38). Different traditions in statistics can be compared to different traditions in contemporary epistemology, because all are rooted in different ways to understand justification: "The key to bridging statistics and epistemology is the concept of *justification*" (54). Thus, Bayesian statistics uses a notion of justification that is similar to that of epistemological internalists, classical statistics hinges on justification in the externalist sense, model selection favors a pragmatist form of justification, and deep learning takes its inspiration from justification in virtue epistemology.

In this commentary, we will focus on the connections between Bayesian statistics and internalist epistemology, which form the subject of Chapter 2. Otsuka endorses the standard objection against both Bayesianism and internalism, viz. that they are in danger of losing contact with the outside world, running the risk of confining inductive reasoning "only to the system of beliefs internal to an epistemic agent and the logical relations between them" (70). For internalism in general, and for some forms of Bayesianism, this is fair enough. However, he also identifies Bayesian updating with internalist justification, and this move, we think, might be too quick.

There seems to be a fundamental difference between justifying and updating. For one thing, while the concept of updating has a clear definition (calculating the posterior probability of a hypothesis from its prior probability by means of Bayes' theorem), the meaning of justification remains notoriously unclear. "[I]t is unlikely that epistemologists will ever agree on what concept of justification is the 'correct' one," writes Otsuka, "or even that the question has a definitive answer" (106). Indeed, and this lack of consensus made epistemologists such as Roderick Chisholm, William Alston, and Richard Swinburne even go so far as to declare that a definition of justification is a chimera (Chisholm 1966, 5–6; Alston 1993, 534; 2005, 22; Swinburne 2001).

But all is not lost. Nowadays, most epistemologists share at least one intuition about justification, namely that it has something to do with probability. Whatever the exact meaning of "A_0 is justified by A_1," this phrase does imply, as a necessary but not a sufficient condition, that A_1 makes A_0 more probable. As a result, the difference between internalists and externalists has now shifted to the meaning of "A_1 makes A_0 more probable." Internalists interpret this expression as that the belief in proposition A_1 will raise the degree of belief in A_0, while for externalists, it means that the presence of event A_1 will increase the frequency of event A_0.

We think that this shift is definitely a step forward in the discussion, since both factions now base themselves on the same probability calculus—only the interpretations differ. This is not to say that interpretations are unimportant. We fully agree with Otsuka when he writes that statistics "cannot cast off the semantic issues of how its mathematical machinery comes to have an empirical significance" (179). At the same time, the important role of mathematics is "a healthy sign of the maturity and soundness of the discipline" (178), and below, we will use the mathematics of probability to explain more clearly the difference between justifying a belief and updating it.

3 Justifying is not updating

Following Rudolf Carnap, we formalize "A_1 makes A_0 more probable" as

$$P(A_0|A_1) > P(A_0)\,, \tag{1}$$

that is, the conditional probability of A_0 given A_1 is greater than the unconditional probability of A_0; in the vocabulary of Carnap, A_0 is confirmed, or rendered firmer, by A_1 (Carnap 1962, xv-xvi). Inequality (1) is not a definition of justification, if only because it is already satisfied when, for example, the value of $P(A_0|A_1)$ is 0.67 while that of $P(A_0)$ is 0.66, in which case we would fight shy of saying that A_1 *justifies* A_0. However, as intimated, a definition of justification will very likely *imply* something like (1). Not without reason the dozen or so different confirmation measures that have been proposed during the past 25 years all presuppose (1). The assumption that (1) is necessary for justification will prove to be enough for our argument.

So far about justification. What about updating? The Bayesian will say that, if proposition A_1 justifies proposition A_0 in the sense of (1), then one should update one's belief in A_0. That is, one should replace the numerical value of $P(A_0)$, the prior, by that of $P(A_0|A_1)$, the posterior. This shows that there is indeed a connection between justifying our belief in A_0 and updating it. However, that there are also significant differences becomes clear when we are dealing with more than two propositions.

To illustrate this, let us begin by adjoining a third proposition, A_2, which justifies A_1, which in turn justifies A_0:

$$P(A_1|A_2) > P(A_1) \qquad\qquad P(A_0|A_1) > P(A_0)\,.$$

The probability of the target proposition A_0 is given by the rule of total probability:

$$P(A_0) = \alpha_0 P(A_1) + \beta_0[1 - P(A_1)]\,,$$

where $\alpha_0 = P(A_0|A_1)$ and $\beta_0 = P(A_0|\neg A_1)$. This is equivalent to the following:

$$P(A_0) = \beta_0 + \gamma_0 P(A_1)\,, \tag{2}$$

where $\gamma_0 = \alpha_0 - \beta_0$. Since A_0 is more probable if A_1 is true than if it is false, γ_0 is positive.

Similarly, the probability of A_1 is

$$P(A_1) = \beta_1 + \gamma_1 P(A_2)\,, \tag{3}$$

where $\alpha_1 = P(A_1|A_2)$, $\beta_1 = P(A_1|\neg A_2)$ and $\gamma_1 = \alpha_1 - \beta_1$. On replacing $P(A_1)$ in Eq. (2) by the right hand side of Eq. (3), we obtain

$$P(A_0) = \beta_0 + \gamma_0(\beta_1 + \gamma_1 P(A_2))\,,$$

or equivalently

$$P(A_0) = \beta_0 + \gamma_0\beta_1 + \gamma_0\gamma_1 P(A_2)\,. \tag{4}$$

Formula (4) expresses the amount of support that A_0 receives from a justificatory chain, ending with the foundational proposition A_2. The formula makes clear that part of the support for A_0 comes from the conditional probabilities α and β (since $\gamma = \alpha - \beta$), and part comes from the unconditional probability $P(A_2)$.

On the Bayesian account, things are quite different. One first updates the value of the probability of A_0 from $P(A_0)$ to that of $P(A_0|A_1)$:

$$P(A_0|A_1) = \frac{P(A_1|A_0)}{P(A_1)} P(A_0)\,;$$

and then one further updates to the value of $P(A_0|A_2)$ by using $P(A_0|A_1)$ as a pseudo-prior for the second updating:

$$P(A_0|A_2) = \frac{P(A_2|A_0)}{P(A_2)} P(A_0|A_1) = \frac{P(A_2|A_0)}{P(A_2)} \frac{P(A_1|A_0)}{P(A_1)} P(A_0) .$$

The result is the same as if one had performed just one update in one fell swoop, from the value of $P(A_0)$ to that of $P(A_0|A_1 \wedge A_2)$[1]:

$$P(A_0|A_2) = \frac{P(A_2 \wedge A_1|A_0)}{P(A_2 \wedge A_1)} P(A_0) . \tag{5}$$

This shows that repeated updating à la Bayes is very different from creating a justificatory chain in epistemology. Although repeated updating may well be temporally sequential, it is not logically so. It does not matter whether you update a belief in A_0 by a belief in A_1 and then use A_2, or whether you update by first using A_2 and then A_1—all evidence is *ex aequo*. In a justificatory chain, however, order is very important. It makes a difference whether you justify A_0 by A_1 and the latter by A_2, or whether A_0 is justified by A_2 which is justified by A_1. Indeed, since probabilistic support is not transitive, A_2 may not justify A_0 at all. In a justificatory chain, evidence is not *ex aequo*, for as we will explain later, the further away in the chain the evidence is, the smaller its role in the justification of A_0.

In Bayesian statistics, prior and posterior constantly change roles in that the new prior takes on the value of the old posterior. In justification, this is not so. Although $P(A_0)$ usually gets a different value as the justificatory chain becomes longer, this new value comes about in a different way and is not identical to some old conditional probability.

4 Washing out vs. fading away

Typical for Bayesian statistics is the use of prior probabilities, which are widely considered to be the Achilles heel of the method. Here is Otuska's example (59): suppose Alice has been tested positive for a certain disease with a test kit that detects 95 of 100 cases and yields a false positive in only 1 out of 10 cases. What is the probability that Alice is actually ill? The answer depends on what we take as our prior probability, in this case, the number of people who have the disease in a particular population at a particular time. Sometimes the prior probability is completely unknown, and then we have to guess. Dependent on what we take as our prior, Alice's chance of having the disease might be one in ten or one in a thousand.

Otsuka recalls the standard reply of the Bayesian to this objection:

> "True, we do not reach an objectively justified conclusion in one shot. Bayesian inference, however, is a process of updating beliefs, and it is by repeating this

[1] This is true if A_1 and A_2 are independent of one another, as is usually the case. If they are not, the formula is more complicated, but the basic point remains unchanged.

> process that we can arrive at the right conclusion. . . . [A]fter a sufficiently large number of n trials, the Bayesian inference will eventually lead us to the same single posterior distribution, regardless of what priors we began with. . . . In general, as we obtain more and more data, the effect of priors gets washed out and the posterior distribution converges to the true parameter . . . Hence, subjective disagreement prior to an inquiry does not pose a serious problem if we have enough data, or so the Bayesians argue." (59–60)

This is the famous "washing out of the prior" that Bayesians use to show that the posterior distribution will approach the truth in repeated updating. In terms of our formalization above, repeated updating implies continuing formula (5) indefinitely, so that we get the following:

$$P(A_0|A_n) = \frac{P(A_n \wedge \ldots \wedge A_3 \wedge A_2 \wedge A_1|A_0)}{P(A_n \wedge \ldots \wedge A_3 \wedge A_2 \wedge A_1)} P(A_0)\,, \tag{6}$$

where $P(A_0|A_n)$ embodies the true distribution as n goes to infinity, no matter what the original prior was.

Otsuka's objection to washing out is that it fails to be a realistic option: we are finite beings that lack the time, data, and resources to engage in updating *ad infinitum* (71). Otsuka is of course right that we are not made for eternity, but one could ask whether endless updating is really required—in most situations, a stop based on pragmatic reasons appears to be enough.

There exists another objection to washing out, not mentioned by Otsuka, which seems to be more telling. The adoption of washing out, as a way to defend the claim that Bayesian statistics is objective, comes at a heavy price. For what one gets in the limit is simply a relative frequency, and the success of repeated Bayesian updating lies precisely in its tending to this limit. Welcoming washing out thus means embracing frequentism, i.e., the classical statistical approach that Otsuka describes in Chapter 3, and connects with epistemological externalism rather than with epistemological internalism. Subjective Bayesians who endorse washing out may therefore be accused of trying to have their cake and eat it too (Atkinson & Peijnenburg 2017, see Appendix C.2 for an example).

This kind of accusation does not apply to an infinite justificatory chain. Suppose our belief in proposition A_0 is justified by our belief in A_1, which is justified by the belief in A_2, and so on, to the belief in A_n. So Eq. (4) is lengthened to the following:

$$P(A_0) = \beta_0 + \gamma_0\beta_1 + \gamma_0\gamma_1\beta_2 + \gamma_0\gamma_1\gamma_2\beta_3 + \ldots + \gamma_0\gamma_1 \ldots \gamma_{n-1} P(A_n)\,. \tag{7}$$

As in formula (4), the justification for A_0 comes partly from the conditional probabilities α and β (recall that $\gamma = \alpha - \beta$) and partly from the unconditional probability $P(A_n)$. The longer the chain is, and thus the greater the distance between the target A_0 and the foundational proposition A_n, the smaller the contribution of A_n to the justification of A_0. If n goes to infinity, the contribution vanishes completely. Elsewhere, we have called this effect "fading foundations" (Atkinson & Peijnenburg, 2017).

Washing out and fading foundations are both regresses in which something goes to zero, but as we have seen, the two regresses differ considerably. The washing out of the prior in repeated updating amounts to replacing the Bayesian procedure by a classical statistical approach, and it can moreover come about in one go. The fading of the foundation in a justificatory chain, by contrast, consists in the nontrivial replacement of the probability of a foundational proposition, A_n, by the conditional probabilities, α_0, β_0, α_1, β_1, and so on, as the major determinants of the probability of the target proposition, A_0.

In line with his identification of update and justification, Otsuka regards the two regresses as being the same. Talking about the justification of a belief by another belief, he writes:

> "One can then easily imagine those beliefs in turn requiring yet other beliefs for their own justification, leading to an infinite regress. We finite beings, however, cannot complete such an infinite chain of justification. Historically, this *regress problem* has presented itself as the first obstacle for internalists in establishing the truth-conduciveness of their concept of justification. ... The same kind of problem may also arise for Bayesianism *qua* internalist epistemology ... we encounter a regress situation here similar to the one that has troubled philosophical epistemology." (58)

And talking about Bayesian updating:

> "This process of accumulating data can be understood as a sort of justificatory 'regress', similar to the one we saw in internalist epistemology (60) ...the posterior distribution will eventually converge to the truth should one repeat the process of justification indefinitely." (61)

Just as updating differs from justification, and washing out differs from fading away, so a prior differs from a foundation. In Bayesian statistics, we start with a prior which we then update. In epistemological justification, however, we do not start with a foundation; rather, we justify A_0 by A_1, which we justify by A_2, and so on, until we reach the foundation where we stop. Otsuka puts prior and foundation on a par:

> "In practice, we never have an infinite amount of data; usually we don't have anything close to that, and in such cases there is a high chance that the prior distribution we took as the premise is not completely washed out ... In such realistic situations where one can afford only a finite chain of justification, one must choose an appropriate prior distribution as the starting point. *The chosen prior is then expected to serve as a base premise, or foundation*, for sustaining the subsequent updating process. In epistemology, such a strategy is known as *epistemological foundationalism*." (61, first italics by us)[2]

5 Justification as trade-off

Otsuka seems to be suggesting that the age-old regress problem in epistemology can be solved by taking into account the washing out procedure in Bayesian statistics. But

[2] In footnote 7 of Chapter 2, Otsuka writes as if Peter Klein's infinitism is about Bayesian updating; in fact, it is about epistemological justification. See also p. 73.

as we have seen, washing out applies to updating, not to justifying beliefs. Moreover, washing out is not needed to solve the problem—fading foundations can do the trick.

Here is how it works (Peijnenburg & Atkinson 2019, 193–194; Atkinson & Peijnenburg 2017, 107–115). Imagine the shortest chain there is: our belief in A_0 is justified by our belief in A_1. Here, A_0 is the target, and A_1 is the foundation; "justified" is interpreted as before, namely as implying, as a necessary but not sufficient condition, that the foundation makes the target more probable. Since the value of $P(A_0)$ lies between those of $P(A_0|A_1)$ and $P(A_0|\neg A_1)$, it cannot be zero if neither of these conditional probabilities is zero.

Suppose $P(A_0|A_1) = x$ and $P(A_0|\neg A_1) = y$, and let x and y differ greatly; for example, x is very close to one, and y is very close to zero. If we do not know $P(A_1)$, we face a great deal of uncertainty as regards the value of $P(A_0)$. The only thing we know is that it lies in the wide interval between x and y. However, fading foundations implies that the interval shrinks as the chain is lengthened. The more beliefs we add, the smaller the interval within which $P(A_0)$ must lie. In the limit of an infinite justificatory chain, this interval has shrunk to a point; the value of $P(A_0)$ has now been determined uniquely in terms of all the conditional probabilities along the chain.

This means that one can specify in advance how many beliefs an agent needs in order to approach the true probability of the target within a given error margin. If the number of beliefs happens to be too large to fit into the agent's finite mind, then she will have to relax the level, and be content with a degree of justification that is less accurate. But if the number is rather small, so that the beliefs are mentally encompassed with ease, then the satisfaction level can always be tightened up and brought closer to the target's true probability.

Epistemic justification thus boils down to a trade-off between the number of beliefs that we can handle and the level of accuracy that we want to reach. If we are unable or unwilling to manage a large number, we have to pay in terms of a lack of precision and hence of trustworthiness of (our belief in) proposition A_0. Taking the short route thus comes at a price, but in situations where precision is not important, we can take it easy and forget about the rest of the chain.

6 Afterword

We have argued that there is a difference between justification and update. Justification we have characterized as probability raising, i.e., as confirmation à la Carnap, with the formula $P(A_0|A_1) > P(A_0)$ as a necessary but not sufficient condition. However, could we not say that update is also justification, albeit of a different kind? (Thanks to an anonymous reviewer for making this point.)

It seems that we could. Updating does not require probability raising, yet even if the posterior, $P(A_0|A_1)$, is less than the prior, $P(A_0)$, an expression like "The posterior is more justified than the prior" sounds not unnatural. On the other hand, in Otsuka's book, no distinction is made between justification as update and justification as confirmation. Otsuka's aim is to connect (internalist) justification with update, and in the rare case that he uses the word "confirmation," he is actually talking about update.[3] As explained above, we think that confirmation and update are not the same, and, moreover, that the traditional description of epistemic internalism and the regress problem is closer to confirmation than to update.

Can we perhaps—following a hunch of the reviewer—defend Otsuka's stance by saying that Bayesian updating makes a better fit with Quinean holism? Although Otsuka himself does not use the argument, he seems to endorse the received view that, according to Quine, it is our entire web of beliefs that faces "the tribunal of sense experience," whereas Carnapian confirmation takes place "in a piecemeal fashion, by picking up a particular belief one by one and checking if it matches with a piece of experience" (68).

We are not sure whether such a defense would work. First, it overlooks the fact that already in the thirties Carnap, too, ventilated holistic views, and second, a Carnapian chain of confirmation can branch out and form a network as well. True, this network does not have the shape of a web with axiomatic sentences in the center and auxiliary ones at the edges; rather, it looks more like a fractal (Atkinson & Peijnenburg, 2012). It is not quite clear how exactly such a fractal fares when faced with the tribunal of sense experience, but does not the same apply to a Quinean web? The latter is after all a metaphor—who can tell what the shape of our body of knowledge is?

An important subject in Otsuka's book, that we have not touched upon at all, concerns the notion of causality. Each of the chapters 2, 3, and 4 skirts around it, but it is in Chapter 5 (and also in the concluding Chapter 6) that Otsuka confronts the subject head-on. Fire causes a fire alarm to go off, and yet the propositions "A fire has started" (f) and "The alarm has gone off" (a) support each other. So $P(a|f) > P(a)$, but also $P(f|a) > P(f)$—the probability calculus does not pick out the cause from the effect.

Otsuka goes into admirable and lucid detail concerning the ways of remedying this shortcoming; in particular, he spells out the lore of Directed Acyclic Graphs (DAGs) and the causal Markov condition. Although this tactic helps to separate cause from effect in cases where there is a collider in the DAG, it does not work in general. What works better, although still not in all cases, is intervention, combined with the **do-**calculus of Judea Pearl: if one suppresses the fire the alarm will no longer sound, but if one instead disconnects the alarm the fire will not thereby be quenched. This illustrates, as Otsuka rightly emphasizes, that more is needed to understand causal connections than reflecting on probability relations or on intervention.

Acknowledgements We thank Masashi Kasaki for the invitation to contribute and an anonymous reviewer for constructive remarks.

[3] See for example the section "Confirmation and Disconfirmation of Hypotheses," which is all about update and in which the word "(dis)confirmation" does not occur.

References

Alston, W. P. (1993). Epistemic desiderata. *Philosophy and Phenomenological Research, 53*, 527–551.

Alston, W. P. (2005). *Beyond 'justification': dimensions of epistemic evaluation*. Ithaca / London: Cornell University Press.

Atkinson, D., & Peijnenburg, J. (2012). Fractal patterns in reasoning. *Notre Dame Journal of Formal Logic, 53*, 15–26.

Atkinson, D., & Peijnenburg, J. (2017). *Fading foundations. Probability and the regress problem.* Cham: Springer. Open access: https://link.springer.com/book/10.1007/978-3-319-58295-5

Carnap, R. (1962). *Logical foundations of probability* (2nd ed.). Chicago: University of Chicago Press.

Chisholm, R. M. (1966). *Theory of knowledge*. Englewood Cliffs, New Jersey: Prentice Hall.

Goodman, N. (1954). *Fact, fiction, and forecast*. Fourth edition, 1983. Cambridge (Mass.): Harvard University Press

Otsuka, J. (2023). *Thinking about statistics. The philosophical foundations*. New York/London: Routledge.

Pearl, J. (2000). *Causality: models, reasoning, and inference*. Cambridge: Cambridge University Press.

Peijnenburg, J., & Atkinson, D. (2019). Finite minds and open minds. In: Branden Fitelson, Rodrigo Borges, and Cherie Braden (eds), *Themes from Klein. Knowledge, scepticism, and justification.* pp. 189–197. Cham: Springer

Swinburne, R. (2001). *Epistemic justification*. Oxford: Oxford University Press.