Pittsburgh Interdisciplinary Mathematics Review **5** (2026), 41–53

Gruesome predictions

Sven Neth

(Communicated by Leonardo Finzi
Copyedited by Vipin Bhat)

Abstract

I explain basic problems of inductive reasoning from a Bayesian point of view. To make probabilistic predictions, we need to start with a prior, but which prior should we choose? I discuss some natural options and the problems they face, with particular attention to 'gruesome' alternatives.

Suppose we are observing emeralds. On the basis of our observations, we would like to predict the color of the next emerald. This is the problem of inductive learning: making inferences from the observed to the unobserved. How should we approach this problem?

We can't be sure about the color of the next emerald. Even if we observe many green emeralds, it is possible that the next emerald turns out to be blue. The best we can hope for are probabilistic predictions. For example, after observing many green emeralds, it might be likely that the next emerald is green.

1. Probability

Let Ω be a finite set. The elements of Ω represent possible *states* and subsets of Ω represent possible *events*. For our purposes, a *probability function* maps each subset of Ω to a real number representing how likely this event is. We interpret probability as degrees of belief or *credences* of a rational agent. This is called the subjectivist or Bayesian interpretation of probability and contrasts with views where probability refers to some objective feature in the world, like relative frequency.

Formally, a probability function $P : \mathcal{P}(\Omega) \to \mathbb{R}$ is defined by:

1. Non-negativity: $P(A) \geq 0$ for all $A \subseteq \Omega$.

2. Normalization: $P(\Omega) = 1$.

3. Finite Additivity: $P(A \cup B) = P(A) + P(B)$ for all $A \subseteq \Omega$ and $B \subseteq \Omega$ with $A \cap B = \varnothing$.

We could generalize this definition by replacing the power-set of Ω with some algebra of subsets of Ω. We could also strengthen finite additivity to countable additivity (Kolmogorov 1933). We set these complications aside and stick with the bare-bones framework just introduced. We also set aside the question whether these axioms are justified. There are other deep philosophical questions about this framework. For example, why are predictions with high probability any better than predictions with low probability if probability is subjective? Instead, we focus on the following question: how can we use probability to make predictions?

We also introduce the concept of conditional probability. If $P(B) > 0$, write $P(A \mid B)$ for the conditional probability of A given B, where:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Intuitively, $P(A \mid B)$ is the probability of A on the supposition that B. We will also think of $P(A \mid B)$ as the probability that a rational agent with credences P will assign to A after learning B and nothing more. So we assume that learning goes via Bayesian conditionalization.

2. **Priors**

Let us now use the Bayesian framework introduced above to formalize our problem. Start with the simplest possible example. We will observe two emeralds which can be either green or blue. We have the following set of possible states:

$$\Omega = \{GG, GB, BG, BB\}.$$

Ultimately, our goal is to make probabilistic predictions, so we care about the probability that the second emerald is green given what we learned about the first emerald. If learning goes via conditionalization, this question can be reduced to the choice of the probability function we start with, which we call the *prior*.

For example, suppose we want to know the conditional probability that the second emerald is green given that the first emerald is green. Using the definition of conditional probability and assuming $P(\text{first green}) > 0$, we have[1]

$$P(\text{second green} \mid \text{first green}) = \frac{P(\text{both green})}{P(\text{first green})} = \frac{P(GG)}{P(GG, GB)}.$$

[1] I write $P(x)$ to abbreviate $P(\{x\})$ and $P(x, y)$ to abbreviate $P(\{x, y\})$.

The conditional probability is already fixed by the prior probability we assign to GG and GB. There is something a bit alien about this. A Bayesian does not make up her mind after seeing the evidence. Instead, she already starts with very specific conditional opinions about what to think given that she receives various bits of evidence. Her policy for making predictions on the basis of new evidence is already written into her prior, fixed from the very beginning of her inquiry.

So, to make predictions, we have to start with some prior P on Ω. But which prior should we choose?

2.1. **Uniformity**

We could choose a uniform prior which assigns the same probability to each possible state, so

$$P(GG) = P(GB) = P(BG) = P(BB) = \frac{1}{4}.$$

This prior is simple. Further, it can be motivated by the natural idea that since we don't know which state obtains, we should treat them all equally and assign the same probability to all of them.

Sadly, the uniform prior does not allow inductive learning. In particular, given that the first emerald is green, the probability that the second emerald is green is $\frac{1}{2}$, which is the same as the prior probability that the second emerald is green. We have:

$$P(\text{second green} \mid \text{first green}) = \frac{P(GG)}{P(GG, GB)} = \frac{1/4}{1/2} = \frac{1}{2} = P(\text{second green}).$$

The uniform prior is equivalent to modeling the colors of our emeralds as independent fair coin flips. This might sometimes be an appropriate model, but it does not allow inductive learning since it makes the colors of different emeralds independent. This argument generalizes to more emeralds. If we start with a uniform prior and observe that 99 out of 100 emeralds are green, the probability that the last emerald is green is still $\frac{1}{2}$. Inductive learning is impossible.

The fact that a uniform prior does not allow inductive learning was first noted by Boole ([1854] 2017, p. 289). It is also discussed by Carnap (1950) and is the key idea behind the No Free Lunch theorem in Machine Learning, which says that under a uniform prior, all learning algorithms have the same expected accuracy (Wolpert 1996; von Luxburg and Schölkopf 2011; Schurz 2021).

2.2. **Frequency indifference**

We are ignorant of how many emeralds are green and blue. So instead of assigning the same probability to all sequences, we could also assign the same probability to all three possible *frequencies* of green and blue emeralds: both green, both blue, one of each. The last event is compatible with two sequences GB and BG and we split probability equally between them. So we end up with the following prior:

$$P(\text{both green}) = P(\text{both blue}) = P(\text{one of each}) = \frac{1}{3}.$$

For the entire space Ω, this prior looks like this:

$$P(GG) = P(BB) = \frac{1}{3},$$

$$P(GB) = P(BG) = \frac{1}{6}.$$

Inductive learning now works as expected. Observing that the first emerald is green updates our probability that the second emerald is green to be greater, since

$$P(\text{second green} \mid \text{first green}) = \frac{P(GG)}{P(GG, GB)} = \frac{1/3}{1/2} = \frac{2}{3} > P(\text{second green}) = \frac{1}{2}.$$

This change in the updated probability is due to assigning a higher probability to GG than to GB. We are privileging uniform sequences of outcomes in the sense of assigning higher probability to them. This kind of prior was endorsed by Bayes (1763) and Carnap (1950) as a model of inductive learning. More recent proponents include Huemer (2009) and Bradley (2020).

2.3. **Gruesome indifference**

Let us say that an emerald is *grue* if it is either green and observed first or blue and observed second (Goodman 1955). So, if our sequence is GB, both emeralds are grue. An emerald is *bleen* if it is either blue and observed first or green and observed second. So, if our sequence is BG, both emeralds are bleen. If our sequence is GG, the first emerald is grue and the second emerald is bleen.

We are ignorant of how many emeralds are grue and bleen. So we could also assign the same probability to all three possible frequencies of grue and bleen. Just as before, if some frequency of grue and bleen emeralds is compatible with more than one sequence, we split the probability equally between them. We end up with the following prior:

$$P(\text{both grue}) = P(\text{both bleen}) = P(\text{one of each}) = \frac{1}{3}.$$

At the level of Ω, this prior looks like this:

$$P(GB) = P(BG) = \frac{1}{3},$$

$$P(GG) = P(BB) = \frac{1}{6}.$$

The first two sequences are assigned a probability of $\frac{1}{3}$ because they correspond to both emeralds being grue or both emeralds being bleen respectively. The last two sequences are assigned a probability of $\frac{1}{6}$ because both of them correspond to one emerald being grue and one emerald being bleen. This prior also seems like a reasonable way to express our uncertainty about the emeralds. Note that this prior does not build in a bias towards green or blue:

$$P(\text{first green}) = P(GG, GB) = \frac{1}{2} = P(\text{first blue}),$$

and analogously for the second emerald.

However, normal inductive learning no longer works. Instead, we get an odd kind of counter-induction: observing that the first emerald is green *decreases* the probability that the second emerald is green:

$$P(\text{second green} \mid \text{first green}) = \frac{1/6}{1/2} = \frac{1}{3} < P(\text{second green}) = \frac{1}{2}.$$

Instead, inductive learning works at the level of grue and bleen. Observing that the first emerald is grue increases the probability that the second emerald is grue:

$$P(\text{second grue} \mid \text{first grue}) = \frac{P(\text{both grue})}{P(\text{first grue})} = \frac{P(GB)}{P(GB, GG)} = \frac{1/3}{1/2} = \frac{2}{3}.$$

The same kind of learning works for bleen. The change in our updated probabilities is due to assigning a higher probability to sequences of outcomes which are uniform in terms of grue and bleen. So which kind of inductive learning is right? The gruesome predictions seem weird to us but try to imagine what things look like from the perspective of people who use grue as a primitive concept. These people will define green in terms of grue and bleen: an emerald is green if it is either grue and observed first or bleen and observed second. They might also find it natural to privilege predictions made in terms of grue and bleen.

This should make us less confident about frequency indifference. There are different ways to 'coarse-grain' sequences, by which I mean to group them together into a partition of events. When we coarse-grain to frequencies of green and blue and assign equal probabilities, we get normal inductive learning. When we coarse-grain to frequencies of grue and bleen and assign equal probabilities, we get gruesome inductive learning. Is there any reason to favor one of them over the other?[2]

[2]Stalker (1994) and Elgin (1997) collect classic papers on the grue problem. Neth (ms) provides a very short introduction.

3. Symmetry

Is there anything we can say to push back against gruesome priors? Before we tackle this question, let us first try to characterize gruesome priors.

The frequency indifference prior embodies two ideas. First, we assign the same probability to each possible frequency of green and blue. Second, if a frequency is compatible with more than one sequence, we split its probability mass equally between all of these sequences.

Let us focus on the second idea. We could also assign different probabilities to different frequencies. In this case, our prior can be represented as some distribution on color frequencies. Which priors are like that? The key idea is that we assign the same probability to sequences which agree on the number of green and blue emeralds but disagree on their ordering. Put differently, order doesn't matter for probability. This property is known as *exchangeability* or *symmetry*, introduced by de Finetti (1937) and independently by Johnson (1924) and Carnap (1950).

To make this idea precise, let us introduce the notion of a *random variable* $X : \Omega \to \mathbb{R}$. Intuitively, a random variable models some quantity that depends on the state of the world and might have different values at different states. We can model our emerald experiment as a sequence of two random variables X_1, X_2, where $X_i = 1$ if the i-th emerald is green and $X_i = 0$ otherwise. More generally, we can model observations of n emeralds as a sequence of random variables $X_1, ..., X_n$.

Now we can define exchangeability relative to a prior P. The sequence of random variables $X_1, ..., X_n$ is exchangeable if $X_1, ..., X_n$ has the same joint distribution as $X_{\pi(1)}, ..., X_{\pi(n)}$ for any permutation π of $\{1, ..., n\}$, where the joint distribution of $X_1, ..., X_n$ is $p_{X_1,...,X_n}(x_1, ..., x_n) = P(X_1 = x_1 \cap ... \cap X_n = x_n)$.[3] Equivalently, $X_1, ..., X_n$ is exchangeable if its joint distribution is a symmetric function of its arguments (Pitman 1993, p. 238).

What properties do exchangeable priors have? Before answering this question, let us think about how to geometrically represent priors. Suppose that there are two outcomes G and B. We can characterize the set of all possible priors as a line between two points: the leftmost point assigns probability one to G and probability zero to B, the rightmost point assigns probability zero to G and probability one to B. All other points are convex combinations of these extreme points. For example, the uniform prior with $P(G) = P(B) = \frac{1}{2}$ is the midpoint of the line.

If we have four outcomes, we also can represent the set of all possible priors geometrically. It is the set of all points (p_1, p_2, p_3, p_4) where $p_i \geq 0$ and $p_1 + p_2 + p_3 + p_4 = 1$. In our case, let us say that $p_1 = P(GG)$, $p_2 = P(GB)$, $p_3 = P(BG)$,

[3] Note that for every $1 \leq i \leq n$, x_i can be any value in the range of X_i.

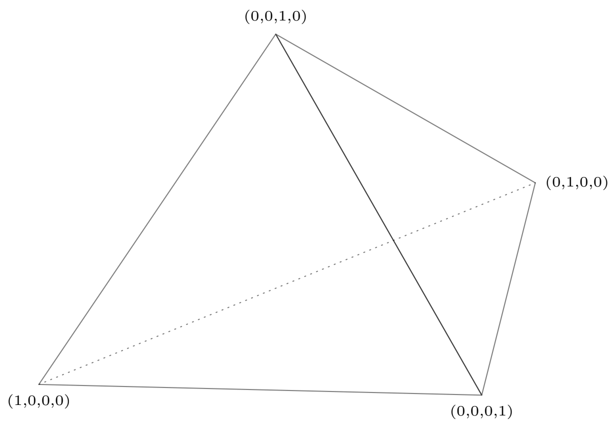

Figure 1: All priors on Ω.

$p_4 = P(BB)$. We can picture this set as a tetrahedron with barycentric coordinate system where the coordinates represent the closeness to the vertices as shown in figure 1 (Diaconis 1977). This geometric representation can be generalized to higher dimensions: all priors on $n+1$ outcomes can be represented as n-dimensional simplex.

The exchangeable priors are given by the plane connecting the points $(1, 0, 0, 0)$, $(0, 0, 0, 1)$ and $(0, \frac{1}{2}, \frac{1}{2}, 0)$ shown in Figure 2. The first point represents the prior which is certain that both emeralds are green, the second point represents the prior which is certain that both emeralds are blue and the third point represents the prior which is certain that one emerald is green and one emerald is blue. All other exchangeable priors are convex combinations of these three extreme points. Note that this includes both the frequency indifference prior and the uniform prior. (Can you picture where they are?) However, it rules out many possible priors, for example, the prior which is certain of GB. This prior violates exchangeability because it assigns probability one to $X_1 = 1$ and $X_2 = 0$ but probability zero to $X_1 = 0$ and $X_2 = 1$.

What does exchangeability do for us? In general, it rules out gruesome indifference. As the perceptive reader might notice, this is not true in the case of two emeralds. Exchangeability is satisfied since $P(GB) = P(BG)$. This means that the gruesome indifference prior in the case of two emeralds has a reasonable interpretation: you are sampling without replacement from an urn with two emeralds and you are $\frac{2}{3}$ confident that one emerald is green and one emerald is blue, with the remaining probability mass split equally between both emeralds being green and both being blue. Diaconis (1977) shows that every finite exchangeable sequence can be represented in this way: as uncertainty over urns that you are sampling from without replacement.

However, if $n > 2$, the gruesome indifference prior violates exchangeability. We

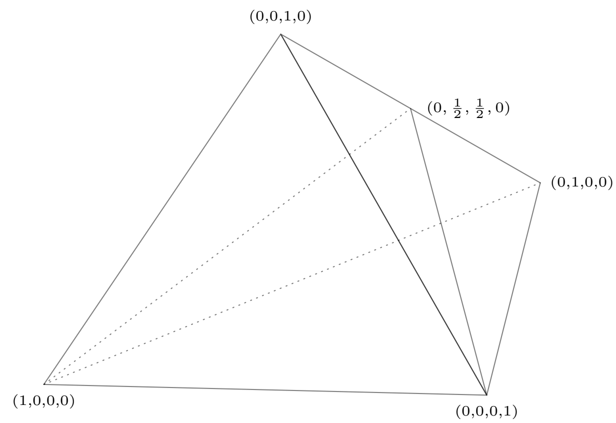

Figure 2: The plane of exchangeability.

can see this by considering the case of three emeralds. The set of states is as follows:

$$\Omega = \{GGG, GGB, GBG, BGG, GBB, BGB, BBG, BBB\}.$$

Let us say that an emerald is grue if it is either observed first and green or observed later (second or third) and blue, and analogously for bleen. In this case, the gruesome indifference prior looks as follows:

$$P(\text{all grue}) = P(\text{two grue, one bleen}) = P(\text{one grue, two bleen}) = P(\text{all bleen}) = \frac{1}{4}.$$

At the level of Ω, this prior looks like this:

$$\text{uniform:} \qquad P(GBB) = P(BGG) = \frac{1}{4},$$

$$\text{two grue, one bleen:} \qquad P(GBG) = P(GGB) = P(BBB) = \frac{1}{12},$$

$$\text{one grue, two bleen:} \qquad P(BGB) = P(BBG) = P(GGG) = \frac{1}{12}.$$

The first two sequences each get $\frac{1}{4}$ of the probability because they correspond to all emeralds being grue and all emeralds being bleen respectively. The second three sequences each get $\frac{1}{12}$ of the probability because they all correspond to two emeralds being grue and one emerald being bleen. In GBG, the first and second emerald are grue and the third emerald is bleen. In GGB, the first emerald is grue, the second emerald is bleen and the third emerald is grue. In BBB, the first emerald is bleen and the second and third emerald are grue. The third three sequences all get $\frac{1}{12}$ of the probability because they all correspond to one emerald being grue and two emeralds being bleen. This prior violates exchangeability since $P(GBB) > P(BBG)$.

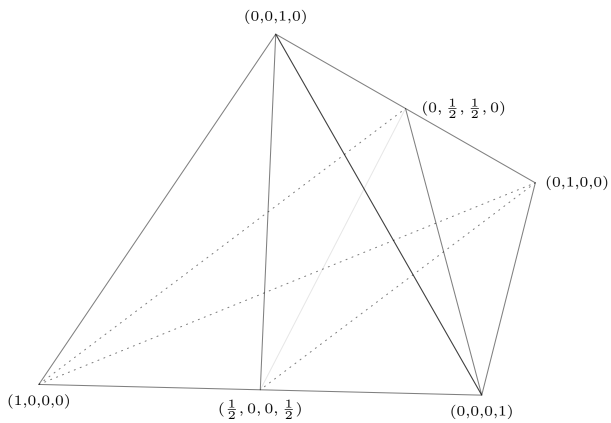

Figure 3: The plane of gruesome exchangeability.

So, exchangeability provides a reason to prefer frequency indifference over gruesome indifference.

However, why accept exchangeability? In particular, why accept exchangeability with respect to green and blue rather than exchangeability with respect to grue and bleen? Above, we described our observations as a sequence of random variables X_1, X_2 where $X_i = 1$ if the i-th emerald is green and $X_i = 0$ otherwise. But we can also describe our observations as a sequence of random variables Y_1, Y_2 where $Y_i = 1$ if the i-th emerald is grue and $Y_i = 0$ otherwise. Why not require that Y_1, Y_2 is exchangeable? This corresponds to assigning the same probability to all sequences which agree on the number of grue and bleen emeralds. Just as green exchangeability allows frequency indifference and rules out gruesome indifference, gruesome exchangeability allows gruesome indifference and rules out frequency indifference.[4]

Geometrically, the plane of gruesome exchangeability is orthogonal to the plane of green exchangeability considered above, as shown in Figure 3. As we can see in the picture, green and gruesome exchangeability are not mutually exclusive but intersect at the line marked in grey. In particular, the uniform prior, which corresponds to the center of mass of the probability simplex, is exchangeable with respect to both green and grue. In recognition of Hume's skepticism about inductive reasoning, we might call this the 'Hume point'.[5]

The priors which satisfy gruesome exchangeability but not green exchangeability are quite strange. A simple example of such a prior assigns probability one to GB but probability zero to BG. According to this prior, it is certain that the first emerald is green and the second emerald is blue. You might think that this prior cannot model

[4]von Kutschera (1972), Earman (1992), and Skyrms (1994) discuss gruesome exchangeability.

[5]Thanks to the students in my spring 2025 grad seminar on grue for suggesting this label.

a situation where you're randomly sampling emeralds with observation-independent colors (Neth forthcoming). If there is one green and one blue emerald and you are randomly sampling, you can observe the emeralds in both possible orderings GB and BG. However, the prior we are currently discussing says that GB has probability one and BG has probability zero so it seems to incorrectly model the inductive situation. Such considerations might give us a reason to reject the requirement of gruesome exchangeability which is independent of which language we speak or which primitives we adopt. Rather, the reason is based on assumptions about our inductive situation: we are randomly sampling and the colors of emeralds are probabilistically independent of when they are observed.

However, perhaps gruesome priors just seem strange because grue is unfamiliar. After all, it isn't so strange to assign a probability of one to GG, which corresponds to being certain that the first emerald is grue and the second emerald is bleen. From a subjectivist point of view, different types of exchangeability correspond to different judgments of how to do inductive reasoning and there is no objective way to break the symmetry (Skyrms 1994). The grue-lovers just roll in a different way and there is no objectively correct way to roll.

4. **Beyond**

Exchangeability constrains priors but still leaves the field wide open. In particular, exchangeability is compatible with dogmatic priors, which assign a probability of one to all emeralds being green. We might want further constraints, for example, a constraint of *regularity* which requires assigning positive probability to each possible sequence. But (green) exchangeability and regularity are still not enough to give us inductive learning, since they leave the possibility of the uniform prior which does not learn anything but treats observations as independent coin flips.

We could think about more constraints. Instead, we'll close by raising a problem for exchangeability. Exchangeable priors are only sensitive to the relative frequency of green and blue emeralds but not to their ordering. Suppose, for example, that we observe the following 'switchy' pattern:

$$GBGBGBGBGB...$$

It seems reasonable that at some point, we should predict the'switchy' pattern to continue. However, exchangeable priors will never do that. For example, the frequency indifference prior will predict that the next emerald is green with probability around $\frac{1}{2}$ after observing the 'switchy' pattern for any length. So exchangeable priors can only do very simple kinds of inductive learning and cannot learn more complicated patterns (Goodman 1946; Good 1969). There are generalizations of exchangeability that can capture some of these patterns, but things get complicated quickly (Diaconis and Freedman 1980).

In general, the question of how to do inductive reasoning remains deeply puzzling. This is so even if we set aside the question why inductive reasoning is possible at all, raised by Hume. Even if our ambitions are modest and we just want to understand what we do when doing inductive reasoning, it is hard to find a clear answer. Still, probability seems like the right framework. And perhaps the most important thing we learn from a probabilistic approach to inductive reasoning is that it is *hard* to make predictions, and we should approach doing so with an open mind. As Hume writes:

> Probability arises from an opposition of contrary chances or causes, by which the mind is not allow'd to fix on either side, but is incessantly tost from one to another, and at one moment is determin'd to consider an object as existent, and at another moment as the contrary. The imagination or understanding, call it which you please, fluctuates betwixt the opposite views; and tho' perhaps it may be oftner turn'd to the one side than the other, 'tis impossible for it, by reason of the opposition of causes or chances, to rest on either. The *pro* and *con* of the question alternately prevail; and the mind, surveying the object in its opposite principles, finds such a contrariety as utterly destroys all certainty and establish'd opinion. (Hume [1739] 1896, p. 440)

References

Bayes, Thomas (1763). "An Essay Towards Solving a Problem in the Doctrine of Chances." In: *Philosophical Transactions of the Royal Society of London* 53. With an introduction by R. Price, pp. 370–418. DOI: 10.1098/rstl.1763.0053.

Boole, George ([1854] 2017). *An Investigation of the Laws of Thought*. Project Gutenberg. URL: https://www.gutenberg.org/ebooks/15114.

Bradley, Darren (2020). "Naturalness as a Constraint on Priors". In: *Mind* 129.513, pp. 179–203. DOI: 10.1093/mind/fzz027.

Carnap, Rudolf (1950). *Logical Foundations of Probability*. University of Chicago Press.

de Finetti, Bruno (1937). "La Prévision: Ses Lois Logiques, Ses Sources Subjectives". In: *Annales de l'Institut Henri Poincaré* 7.1, pp. 1–68.

Diaconis, Persi (1977). "Finite Forms of de Finetti's Theorem on Exchangeability". In: *Synthese* 36 (2), pp. 271–281. DOI: 10.1007/bf00486116.

Diaconis, Persi and David Freedman (1980). "De Finetti's Generalizations of Exchangeability". In: *Studies in Inductive Logic and Probability*. Ed. by Richard Jeffrey. Vol. II. University of California Press, pp. 233–50. DOI: 10.1525/9780520318328-007.

Earman, John (1992). *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. MIT Press.

Elgin, Catherine Z. (1997). *Nelson Goodman's New Riddle of Induction*. Vol. 2. The Philosophy of Nelson Goodman. Taylor & Francis.

Good, Irving John (1969). "Discussion of Bruno de Finetti's Paper 'Initial Probabilities: A Prerequisite for any Valid Induction'". In: *Synthese* 20.1, pp. 17–24. DOI: 10.1007/bf00567233.

Goodman, Nelson (1946). "A Query on Confirmation". In: *Journal of Philosophy* 43.14, pp. 383–385. DOI: 10.2307/2020332.

— (1955). *Fact, Fiction, and Forecast*. Harvard University Press.

Huemer, Michael (2009). "Explanationist Aid for the Theory of Inductive Logic". In: *British Journal for the Philosophy of Science* 60.2, pp. 345–375. DOI: 10.1093/bjps/axp008.

Hume, David ([1739] 1896). *A Treatise of Human Nature*. Ed. by L. A. Selby-Bigge. Oxford University Press.

Johnson, William Ernest (1924). *Logic, Part III: The Logical Foundations of Science*. Cambridge University Press.

Kolmogorov, Andrey (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer.

Neth, Sven (forthcoming). "Random Emeralds". In: *Philosophical Quarterly*. DOI: 10.1093/pq/pqaf038.

— (ms). "Goodman's New Riddle of Induction Explained in Words of One Syllable". URL: https://philpapers.org/rec/NETGNR.

Pitman, Jim (1993). *Probability*. Springer.

Schurz, Gerhard (2021). "The No Free Lunch Theorem: Bad News for (White's Account of) the Problem of Induction". In: *Episteme* 18.1, pp. 31–45. DOI: 10.1017/epi.2018.54.

Skyrms, Brian (1994). "Bayesian Projectibility". In: *Grue! The New Riddle of Induction*. Ed. by Douglas Stalker. Open Court, pp. 241–62.

Stalker, Douglas Frank (1994). *Grue! The New Riddle of Induction*. Open Court.

von Kutschera, Franz (1972). *Wissenschaftstheorie I: Grundzüge der allgemeinen Methodologie der empirischen Wissenschaften*. Wilhelm Fink Verlag.

von Luxburg, Ulrike and Bernhard Schölkopf (2011). "Statistical Learning Theory: Models, Concepts, and Results". In: *Handbook of the History of Logic*. Ed. by Dov

Gabbay, Stephan Hartmann, and John Woods. Vol. 10. Elsevier, pp. 651–706. DOI: 10.1016/B978-0-444-52936-7.50016-1.

Wolpert, David H. (1996). "The Lack of A Priori Distinctions Between Learning Algorithms". In: *Neural Computation* 8.7, pp. 1341–1390. DOI: 10.1162/neco.1996.8.7.1341.

Sven Neth
Department of Philosophy, University of Pittsburgh, Pittsburgh, PA
Email: sven.neth@pitt.edu