CrossMark

Episteme, 15, 2 (2018) 202–208 © Cambridge University Press
doi:10.1017/epi.2017.2

FOOL ME ONCE: CAN INDIFFERENCE VINDICATE INDUCTION?

ZACH BARNETT AND HAN LI
zachary_barnett@brown.edu, han_li@brown.edu

ABSTRACT

Roger White (2015) sketches an ingenious new solution to the problem of induction. He argues from the principle of indifference for the conclusion that the world is more likely to be induction-friendly than induction-unfriendly. But there is reason to be skeptical about the proposed indifference-based vindication of induction. It can be shown that, in the crucial test cases White concentrates on, the assumption of indifference renders induction no more accurate than random guessing. After discussing this result, the paper explains why the indifference-based argument seemed so compelling, despite ultimately being unsound.

Roger White (2015) sketches an ingenious new solution to the problem of induction, arguing on *a priori* grounds that the world is more likely to be induction-friendly than induction-unfriendly. The argument relies primarily on the principle of indifference, and, somewhat surprisingly, assumes little else. If inductive methods could be vindicated in anything like this way, it would be quite a groundbreaking result. But there are grounds for pessimism about the argument. It can be shown that, in the crucial test cases White concentrates on, the assumption of indifference renders induction no more accurate than random guessing. After discussing this result, we then explain why the indifference-based argument seemed so compelling, despite ultimately being unsound.

1. AN INDIFFERENCE-BASED STRATEGY

White begins by imagining that we are "apprentice demons" tasked with devising an *induction-unfriendly world* – a world where inductive methods tend to be unreliable. To simplify, we imagine that there is a single binary variable that we control (such as whether the sun rises over a series of consecutive days). So, in essence, the task is to construct a binary sequence such that – if the sequence were revealed one bit at a time – an inductive reasoner would fare poorly at predicting its future bits. This task, it turns out, is surprisingly difficult. To see this, it will be instructive to consider several possible strategies for constructing a sequence that would frustrate an ideal inductive predictor.

Immediately, it is clear that we should avoid uniformly patterned sequences, such as:

```
0000000000000000000000000000000
```

or

```
0101010101010101010101010101010
```


https://doi.org/10.1017/epi.2017.2 Published online by Cambridge University Press

Sequences like these are quite kind to induction. Our inductive reasoner would quickly latch onto the obvious patterns these sequences exhibit. A more promising approach, it might seem, is to build an apparently patternless sequence:

0010101001111100001110001001010 0

But, importantly, while induction will not be particularly *reliable* at predicting the terms of this sequence, it will not be particularly *unreliable* here either. Induction would simply be silent about what a sequence like this contains. As White puts it, "In order for . . . induction to be applied, our data must contain a salient regularity of a reasonable length" (p. 285). When no pattern whatsoever can be discerned, presumably, induction is silent. (We will assume that the inductive predictor is permitted to suspend judgment whenever she wishes.) The original aim was not to produce an induction-neutral sequence, but to produce a sequence that elicits errors from induction. So an entirely patternless sequence will not suffice. Instead, the induction-unfriendly sequence will have to be more devious, building up seeming patterns and then violating them. As a first pass, we can try this:

0000000000000000000000000000001

Of course, this precise sequence is relatively friendly to induction. While our inductive predictor will undoubtedly botch her prediction of the final bit, it is clear that she will be able to amass a long string of successes prior to that point. So, on balance, the above sequence is quite kind to induction – though not maximally so.

In order to render induction unreliable, we will need to elicit more errors than correct predictions. We might try to achieve this as follows:

0000111100001111000011110000 1111

The idea here is to offer up just enough of a pattern to warrant an inductive prediction, before pulling the rug out – and then to repeat the same trick again and again. Of course, this precise sequence would not necessarily be the way to render induction unreliable: For, even if we did manage to elicit an error or two from our inductive predictor early on, it seems clear that she would eventually catch on to the exceptionless higher-order pattern governing the behavior of the sequence.

The upshot of these observations is not that constructing an induction-unfriendly sequence is impossible. As White points out, constructing such a sequence should be possible, given any complete description of how exactly induction works (p. 287). Nonetheless, even if there are a few special sequences that can frustrate induction, it seems clear that such sequences are fairly few and far between. In contrast, it is obviously very easy to *corroborate* induction (i.e. to construct a sequence rendering it thoroughly reliable). So induction is relatively *un-frustrate-able*. And it is worth noting that this property is fairly specific to induction. For example, consider an inferential method based on the gambler's fallacy, which advises one to predict whichever outcome has occurred less often, overall. It would be quite easy to frustrate this method thoroughly (e.g. '00000000. . .').

So far, we have identified a highly suggestive feature of induction. To put things roughly, it can seem that:

https://doi.org/10.1017/epi.2017.2 Published online by Cambridge University Press

* Over a large number of sequences, induction is thoroughly reliable.
* Over a large number of sequences, induction is silent (and hence, neither reliable nor unreliable).
* Over a very small number of sequences (i.e. those specifically designed to thwart induction), induction is unreliable (though, even in these cases, induction is still silent much of the time).

Viewed from this angle, it can seem reasonable to conclude that there are *a priori* grounds for confidence that an arbitrary sequence is not induction-unfriendly. After all, there seem to be far *more* induction-friendly sequences than induction-unfriendly ones. If we assign equal probability to every possible sequence, then the probability that an arbitrary sequence will be induction-friendly is going to be significantly higher than the probability that it will be induction-unfriendly. So a simple appeal to the principle of indifference seems to generate the happy verdict that induction can be expected to be more reliable than not, at least in the case of binary sequences.

Moreover, as White points out, the general strategy is not limited to binary sequences. If we can show *a priori* that induction over a binary sequence is unlikely to be induction-unfriendly, then it's plausible that a similar kind of argument can be used to show that we are justified in assuming that an arbitrary *world* is not induction-unfriendly. If true, this would serve to fully vindicate induction.

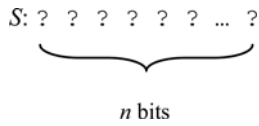
2. GIVEN INDIFFERENCE, INDUCTION IS NOT RELIABLE

However, there are grounds for pessimism about whether the strategy is successful – even in the simple case of binary sequences. Suppose that, as a special promotion, a casino decided to offer *Fair Roulette*. The game involves betting $1 on a particular color – black or red – and then spinning a wheel, which is entirely half red and half black. If wrong, you lose your dollar; if right, you get your dollar back and gain another. If it were really true that induction can be expected to be more reliable than not over binary sequences, it would seem to follow that induction can serve as a winning strategy, over the long term, in Fair Roulette. After all, multiple spins of the wheel produce a binary sequence of reds and blacks. And all possible sequences are equally probable. Of course, induction cannot be used to win at Fair Roulette – past occurrences of red, for example, are not evidence that the next spin is more likely to be red. This suggests that something is amiss. Indeed, it turns out that no inferential method – whether inductive or otherwise – can possibly be expected to be reliable at predicting unseen bits of a binary sequence, if the principle of indifference is assumed.[1]

To illustrate this: Let S be an unknown binary sequence of length n. S is to be revealed one bit at a time, starting with the first.

1 This fact is a direct consequence of Wolpert's (1996: 1354) "No Free Lunch" theorem, which, among other things, places limits on the expected accuracy of computable predictive functions, given certain assumptions about the relative probabilities of different occurrences and given certain other assumptions about how predictive functions are scored. The theorem and its proof are complicated. Here, we provide a relatively straightforward argument for a claim that turns out to be an immediate corollary. See Schurz (Ms.) for a discussion of how Wolpert's results bear on the problem of induction more generally. Thanks to the editors of *Episteme* for making this connection.

https://doi.org/10.1017/epi.2017.2 Published online by Cambridge University Press



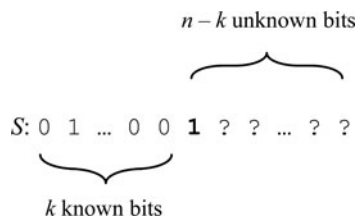
Let f be an arbitrary predictive function that takes as input any initial subsequence of S and outputs a prediction for the next bit: '0', '1', or 'suspend judgment'.

A predictive function's accuracy is measured as follows: +1 for each correct prediction; –1 for each incorrect prediction; 0 each time 'suspend judgment' occurs. (So the maximum accuracy of a function is n; the minimum score is $-n$.) Given a probability distribution over all possible sequences, the *expected accuracy* of a predictive function is the average of its possible scores weighted by their respective probabilities.

We now arrive at an important fact: If we assume indifference (i.e. if we assign equal probability to every possible sequence), then – no matter what S is – each of f's predictions will be expected to contribute 0 to f's accuracy; and as a result, f has 0 expected accuracy, generally.

This fact can be shown as follows. For some initial subsequences, f will output 'suspend judgment'. The contribution of such predictions will inevitably be 0. So we need consider only those cases where f makes a firm prediction (i.e. '0' or '1'; not 'suspend judgment').

Let K be a k-length initial subsequence for which f makes a firm prediction about the bit in position $k + 1$. Specifically, suppose that f predicts that '1' will be in position $k + 1$.



Consider the full sequences that begin with K and for which the prediction is correct. These are the sequences that begin with K and have '1' in position $k + 1$. There are $2^{n-(k+1)}$ of these sequences, since there are $2^{n-(k+1)}$ ways that this sequence could terminate. But there are also exactly $2^{n-(k+1)}$ sequences beginning with K where '1' is *not* in position $k + 1$. (For these sequences, '0' is in position $k + 1$ instead.)

So the number of possible sequences that make the prediction correct is equal to the number that make it incorrect. Given indifference, the probability of a correct prediction and the probability of an incorrect prediction both equal .5, which makes the expected contribution of this prediction 0.

Of course, the same reasoning would apply equally if f's prediction were '0' instead of '1'. Indeed, the reasoning generalizes to all of f's predictions. So the expected contribution of every prediction is 0. It follows immediately that f's expected accuracy is 0. The upshot is that if indifference is assumed, then there is absolutely no method, inductive or otherwise, for predicting the unseen bits of a binary sequence that can be expected to perform reliably. In fact, the principle of indifference actually *precludes* induction from being expectedly accurate.

https://doi.org/10.1017/epi.2017.2 Published online by Cambridge University Press

3. A DIAGNOSIS

We have seen that the indifference-based strategy does not work for binary sequences. What, then, is so attractive about it? At least intuitively, it seems right to claim that it is difficult to construct a binary sequence on which induction is consistently unreliable. At best, we can construct sequences on which induction rarely hazards any guesses at all, only occasionally issuing false predictions. But even these are hard to imagine. On the other hand, we saw that it is easy to construct sequences on which induction is wildly successful. How can these observations be squared with the result from §2?

The answer has to do with the nature of the inductive method. Induction takes its own past record of success and failure as evidence for future predictions. If the past has been unkind to induction, then induction will be loath to make further predictions. Confronted with its own past failures, induction is unwilling to stick its neck out again – in this sense, we might say that induction is *shy*. This explains why it is so hard to find binary sequences on which induction is consistently unreliable. Once induction begins to exhibit unreliability, it will stop making predictions at all. On the other hand, induction is especially willing to continue making predictions in the face of past success. Thus, it is easy to construct the sequences on which induction is consistently reliable.

Shyness, however, is not a property that is unique to inductive prediction. And, crucially, shyness is in no way evidence of the reliability of a predictive method. To illustrate, consider the following predictive method:

> **Fool Me Once (FMO)**: Continue predicting '0' until '1' occurs. Then suspend judgment for all subsequent bits.

FMO is quite shy – one of the shyest methods possible. As long as its predictions continue to be confirmed, it will continue to recommend firm predictions. But as soon as it issues a single false prediction, it forever retires from the game, staying silent for the rest of the sequence no matter what happens.

Importantly, FMO has the very same characteristics that the indifference-based strategy relied upon in the case of induction. To see this, note that it is not possible to construct an FMO-unfriendly sequence – a sequence that renders FMO consistently unreliable. At most, we can elicit one false prediction and no true ones. On the other hand, it is easy to construct sequences that render FMO very successful: Any sequence that begins with a long string of '0's will ensure that FMO ends up with a relatively high accuracy score.

So, as with induction, it is in some sense easier to construct a FMO-friendly sequence than a FMO-unfriendly sequence. This suggests that this shyness is the feature of induction the indifference-based strategy relied upon. After all, shyness is the defining characteristic – and, perhaps, the only characteristic – of FMO as a predictive method. It takes shyness to the extreme – even a single false prediction is indefeasible reason to give up making predictions all together – and does nothing else. The mere fact that a predictive method is shy, however, gives us no reason to expect the method to be reliable – at least, if indifference is assumed. Of course, this is a consequence of the result shown in §2 – since no methods can be expected to be reliable whatsoever. But it may be helpful to see why FMO turns out not to be reliable. Doing so will help to illustrate what was so appealing about the indifference-based argument.

https://doi.org/10.1017/epi.2017.2 Published online by Cambridge University Press

Consider an unknown binary sequence of length *n*. FMO continues making predictions until the first '1' occurs, at which point, FMO falls silent. To begin, consider those sequences that begin with '1'. Since these cases comprise half of all possible sequences, the probability of such a sequence's occurrence is .5 (via indifference). In these cases, FMO's score will be −1. Next, consider those sequences that have an initial '0' followed by a '1'. The probability of such a sequence's occurrence is .25, and in these cases FMO's score will be 0. Consider those sequences that begin with two '0's, followed by a '1'. The probability of such a sequence's occurrence is .125, and in these cases FMO's score will be −1.

A pattern emerges. FMO's expected accuracy will be:[2]

$$(.5)(-1) + (.25)(0) + (.125)(+1) + (.06125)(+2) + \ldots + \left(\frac{1}{2^n}\right)(n-2) + \left(\frac{1}{2^n}\right)(n)$$

Ultimately, FMO's expected accuracy on S is:

$$\sum_{k=1}^{n}\left(\frac{1}{2^k}\right)(k-2) + \left(\frac{1}{2^n}\right)(n) = 0$$

Here we can see what is wrong with the indifference-based argument. Though there are no possible sequences on which FMO is consistently unreliable, there are a large number of sequences on which FMO is ever-so-slightly unreliable – indeed, these sequences comprise half of all possible sequences. These cases balance out the comparatively few sequences on which FMO is reliable – including the small number on which FMO is highly reliable.

An analogous point seems to hold for induction, although the details will depend on the specific predictive rule that is taken to constitute inductive reasoning. For illustrative purposes, consider the simple rule which always predicts that the next bit will be *whichever digit has occurred most often* – ignoring where in the sequence those digits have occurred (e.g. so if the sequence so far is '0001', then our rule predicts that '0' will occur next). When the sequence contains equally many occurrences of '0' and '1', no prediction is made. This rule, it should be noted, has much in common with induction: It relies on a version of the assumption that what has happened will continue to happen. And, like induction, it is somewhat shy: enough mistakes will make the rule fall silent (since each mistake brings the sequence closer to a perfect balance of '0's and '1's), while a string of success will ensure that the rule continues to make predictions.

Suppose that a sequence of only five bits is to be revealed one bit at a time. And suppose we use our simple rule to try to predict each bit. First, we should note that there are a few sequences that make our rule massively successful: The sequences '00000' and '11111' both lead to the very high score of +4 for our rule. And we should note that there are no sequences that frustrate our rule to the same degree: −2 is the lowest possible score our rule can ever earn. But in terms of quantity, there are *more* unfriendly sequences than friendly ones, by a count of 16 to 12. As was the case with FMO, the quantity of (somewhat) unfriendly sequences makes up for the discrepancy between the best case and the worst cases. Table 1 summarizes the possible outcomes.

2 Note the last term. This is for the sequence composed exclusively of '0's, since in this case no false predictions are made. In this case, FMO has an accuracy score of *n*.

https://doi.org/10.1017/epi.2017.2 Published online by Cambridge University Press

Table 1 Possible outcomes.

SCORE	COUNT	LIST OF SEQUENCES
+4	2	00000, 11111
+2	6	00001, 00010, 01111, 10000, 11101, 11110
+1	4	01000, 01111, 10000, 10111
0	4	00011, 00101, 11010, 11100
−1	8	00110, 00111, 01001, 01110, 10001, 10110, 11000, 11001
−2	8	01010, 01011, 01100, 01101, 10010, 10011, 10100, 10101

So, where has the indifference-based argument gone wrong? The argument rests on the claims that there were many sequences friendly to induction, and comparatively few that were unfriendly. There are two ways to interpret these claims. If a sequence is "friendly" to induction just in case it elicits predictions from induction that would earn a positive score (with "unfriendly" defined correspondingly), then the claims are simply false. But, more interestingly, if a sequence is "friendly" to induction just in case it elicits predictions from induction that would earn a *very high* positive score (with "unfriendly" defined correspondingly), then the claims are true – but irrelevant to the expected accuracy of induction. For any shy method will have this property, and many shy methods are not particularly reliable.

4. CONCLUSION

Many proposed solutions to the problem of induction require us to have some *a priori* basis for assuming that the world is uniform in a way that makes it amenable to induction. Such solutions are somewhat unsatisfying, as it can be mysterious how we are warranted in making such assumptions. An *a priori* vindication of inductive methods that relies solely on the principle of indifference would address this worry, insofar as the principle more naturally admits of *a priori* justification. We have seen, however, that the indifference-based vindication of induction mistakes shyness for accuracy. So it may be that, in our effort to solve the problem of induction, we are stuck with a certain degree of mystery.[3]

REFERENCES

Schurz, G. Ms. 'No Free Lunch Theorem, Inductive Skepticism, and the Optimality of Meta-Induction.'

White, R. 2015. 'The Problem of the Problem of Induction.' *Episteme*, 12(2): 275–90.

Wolpert, D. 1996. 'The Lack of A Priori Distinctions between Learning Algorithms.' *Neural Computation*, 8: 1341–90.

ZACH BARNETT and HAN LI are graduate students at Brown University whose research interests include traditional and formal epistemology.

3 For helpful comments and suggestions, we thank David Christensen and Josh Schechter, as well as members of the Brown Epistemology Reading Group.

https://doi.org/10.1017/epi.2017.2 Published online by Cambridge University Press