

# **Inference to the Best Explanation and the Principle of Sufficient Reason**

**Jacob Mills**

Houston City College

Winner of the NMTX Award

In the following I argue that underlying our deductive and inductive inferential practices are certain inferential presuppositions. In the case of induction, this is the regularity principle while in the case of deduction, they are whatever basic rules or axioms lie at the foundation of one's first-order deductive system. I believe the existence of these inferential presuppositions are indirectly indicated by Hume's skepticism about induction and Carroll's paradox of deduction. I then turn to a consideration of inference to the best explanation (IBE). I conclude that the inferential presupposition operating behind IBE is the intelligibility principle—if something is true there is a best explanation for it. This intelligibility principle is one form of the Principle of Sufficient Reason (PSR). I then argue that because the contrapositive of this principle is the presupposition underlying PSR inferences, inferences that run from the lack of an explanation of P to P being false, that if one is committed to endorsing IBE inferences one is logically committed to endorsing PSR inferences. Metaphysical consequences follow.

## **CIRCULARITY AND INDUCTIVE PRACTICES**

Hume is known for, amongst other things, his skepticism about inductive inferences. He argues that when we question our enumerative inductive inferences, we see that they enthymatically assume a principle about the regularity of nature. He then questions the status of this regularity principle. It is not analytically true, and thus he says is not known to be true a priori by reason. Furthermore, to assume that it could be known a posteriori via inductive methods would be circular—we

know that nature is regular because it always has been up till now! Thus, all our inductive inferences are irrational or circular. Hume has some additional steps, but I think my summary captures the essentials of his argument. Here is Hume in the *Enquires*:

When it is asked, *What is the nature of all our reasonings concerning matter of fact?* the proper answer seems to be, that they are founded on the relation of cause and effect. When again it is asked, *What is the foundation of all our reasonings and conclusions concerning that relation?* it may be replied in one word, Experience. But if we still carry on our sifting humor, and ask, *What is the foundation of all conclusions from experience?* this implies a new question, which may be of more difficult solution and explication. (32)

The answer to the new question is, of course, the regularity principle. He says it is not grounded in reason because it is synthetic and hence contingent, and hence not known a priori. Again here are Hume's own words, "That there are no demonstrative arguments in the case seems evident; since it implies no contradiction that the course of nature may change..." (35). Furthermore, to attempt to ground it inductively would result in a circle:

To endeavour, therefore, the proof of this last supposition [the regularity principle] by probable arguments, or arguments regarding existence, must evidently be going in a circle, and taking that for granted, which is the very point in question. (35-6)

As is so often the case with Hume, there is much buried insight. After all, it took Hume to awaken Kant from his dogmatic slumbers. First I want to point out that Hume does *not* think his observations about inductive inferences makes them unjustified. He says,

I shall allow, if you please, that the one proposition may *justly* be inferred from the other: I know, in fact, that it is always inferred. But if you insist that the inference is made by a chain of reasoning, I desire you to produce that reasoning. (34; my italics)

Thus Hume appears to think our inductive practices are a species of non-rational and yet epistemically justified inference! Furthermore, for all its imposing appearance, there is much trickery in Hume's skepticism. After all, Hume knew Aristotle well enough to remember what he says in the *Posterior Analytics*:

...there will not be understanding of the principles; and since it is not

possible for anything to be truer than understanding, except comprehension, there will be comprehension of the principles—both if we inquire from these facts and because demonstration is not a principle of demonstration so that understanding is not a principle of understanding either—so if we have no other true kind apart from understanding, comprehension will be the principle of understanding. And the principle will be of the principle, and understanding as a whole will be similarly related to the whole object. (59)

While this passage from the *Posterior Analytics* is notoriously enigmatic, I believe it points in the direction of Hume’s skeptical trick. Aristotle says that if we ask for the principles or foundation of the understanding that is produced by demonstration (what he calls scientific deduction), that principle cannot possibly be produced by demonstration itself! The knowing that is understanding cannot ground itself! And yet this is precisely what Hume does when he challenges us to produce the foundation for inductive inference, saying that it cannot be produced deductively, for then it wouldn’t be inductive, nor can it be produced inductively, because that would be circular.

### INFINITE REGRESS AND DEDUCTIVE PRACTICES

Something very similar is got up to in Lewis Carroll’s “What the Tortoise Said to Achilles.” In this paper Carroll offers a challenge to the validity of deductive inferences deeply connected with Hume’s challenge to inductive ones. The tortoise asks Achilles about the foundation for the simple deductive inference *modus ponens*<sup>1</sup>:

$$\begin{array}{l} 1) P \supset Q \\ 2) P \\ \hline Q \end{array}$$

He asks Achilles how the conclusion Q follows from 1) and 2). Achilles answers that according to the valid inference rule *modus ponens*, from 1) and 2) one can infer Q. The tortoise then responds by saying that technically speaking, the conclusion follows from neither 1) or 2) by themselves, but in combination with the hypothetical “if 1) and 2) then Q,” so that a new premise must be added to the inference expressing this, such as 3):

$$\begin{array}{l} 1) P \supset Q \\ 2) P \\ 3) [(P \supset Q) \ \& \ P] \supset Q \\ \hline Q \end{array}$$

The tortoise continues, claiming that the conclusion *Q* *still* doesn't follow from 1) and 2) and 3) by themselves, but only with the addition of the further hypothetical that "if 1) and 2) and 3) then *Q*." So that we continue needing to add premises to our deductive argument, and we are off to the races.

- 1)  $P \supset Q$
- 2)  $P$
- 3)  $[(P \supset Q) \ \& \ P] \supset Q$
- 4)  $[\{(P \supset Q) \ \& \ P\} \ \& \ (P \supset Q) \ \& \ P] \supset Q$
- ...

---

- $Q$

Now of course, Lewis is engaged in as much jugglery here as Hume was, and of a similar kind. The sentence  $[(P \supset Q) \ \& \ P] \supset Q$  is an expression in the object language of the entailment  $P \supset Q$ ,  $P \vDash Q$ , which is an expression in the meta-language. The latter fact about validity no more needs to be included as a specific premise within an example of modus ponens than an explicit declaration of the regularity principle needs to be included in a given enumerative inference. And just as circularity was the ultimate result of Hume's argument, here, an infinite regress is the result of Carroll's. In the same way that we say that our inductive practices involved the regularity principle of nature as an inferential presupposition, here we see that our deductive practices include specific inference rules as their own inferential presuppositions. Thus, modus ponens is itself the inferential presupposition for any instance of modus ponens.<sup>2</sup> The parallels with Hume's inductive skepticism are clear. Carroll challenges us to produce the foundation for deductive inference, obviously it cannot be produced inductively, for then it wouldn't be deductive, nor can it be produced deductively, because that would lead to an infinite regress.

### **INFERENCE TO THE BEST EXPLANATION AND THE INTELLIGIBILITY PRINCIPLE**

Let us examine the structure of an Inference to the Best Explanation (IBE). Take the following example:

- 1) The porridge I left on the mantle is missing  
A nisse must have eaten it

This is a fairly weak example of IBE, but an example nevertheless. The structure for IBE inferences is the following:

1) Phenomenon P (explanandum)

2)  $\{E_1 \vee E_2 \vee E_3 \vee \dots \vee E_n\}$  (set of possible explanans)

$E_i$  where  $E_i$  is the best

That is, there is a phenomenon P that needs explaining (explanandum). Unsaid and unspoken in most IBE is premise 2), a list of possible explainers for 1) (explanans). The inference rule then is that we are to pick the explainer from the set of possible explainers that is the best. Like all inductive arguments, IBE's can be assessed in terms of strength and cogency.<sup>3</sup> A strong inductive inference is one such that *if* the premises are all true then the conclusion is most likely true. While a cogent inference is both strong and has all true premises. Thus the example above is a weak IBE because a nisse, which is a Scandinavian elf creature, eating the porridge is presumably not the best explanation from the set of possible explainers. In general this is how IBE inferences are weak, due to a failure to pick out the best explanation from the set of possible explainers. Cogency is trickier. An IBE fails to be cogent when, although it is strong, one's set of possible explainers fails to be complete and excludes the best explanation. Then, although one may pick out the relative best explanation, relative to your incomplete set of possible explainers, one still fails to pick out the best explanation, again, because your set of possible explainers is incomplete and fails to include some explanation such that, had it been included it would have beaten the others as the best.

What then is the inferential presupposition underlying IBE? A brief consideration shows that it is not the regularity principle that underlies enumerative and analogical inductive inferences.<sup>4</sup> Someone who engages in and endorses IBE as an inferential practice must presuppose the intelligibility principle—if  $P=T$  then there is a best explanation for P.<sup>5</sup> Without this inferential presupposition the entire practice of IBE simply fails to be intelligible.

### WHEN PROBABILISTIC INFERENCES FAIL

The inferential presuppositions associated with probabilistic inferences are not themselves probabilistic. This is clear in the case of cogent enumerative inferences whose conclusions nevertheless turn out false. When this is the case, we do not think of it as a failure of the regularity principle. The probabilistic nature of the inference is not due to the fact that the regularity principle itself is probabilistic—that nature acts regularly most of the time. Rather, it is due to the complexity of isolating the causal mechanism on the back of which the inference rides. The failure of a cogent inductive inference is due to the fact that we have identified a case of correlation that is not causation!

Similarly in the case of cogent IBE that fail. The failure is not due to a lapse in the intelligibility principle that underlies those inferences. Rather, that failure must be located somewhere in the inference itself. Because the IBE is cogent that

means that not only have we successfully picked out the best explanation relative to our set of possible explainers, but that that set is itself complete, containing all relevant explainers. The failure here must be due to an incompleteness in our concept of bestness, along with general vagaries involved in weighing the different dimensions that make up our concept of bestness.<sup>6</sup> Thus the failure of cogent IBE is due to the fact that our notion of what bestness consists in is both incomplete, vague, and evolving. As we experience the world we constantly try to tailor our notion of intelligibility to the intelligibility that is the “whatness” of the world.<sup>7</sup>

### IBE AND COMMITMENT TO PSR INFERENCES

Having established that the regularity and intelligibility principles that underlie our probabilistic inferences are themselves not probabilistic, we can understand that anyone committed to the intelligibility principle:

IP: If  $P=T$  than there is a best explanation for P

is also committed to its contrapositive:

CIP: If there is no best explanation for P then  $P=F$

We can call this the contrapositive intelligibility principle. Now, CIP is the underlying inferential presupposition behind PSR inferences. These are what Della Rocca calls “explicability arguments,” where it is claimed that something is not the case because there would be no explanation for it if it were (306). These PSR inferences assume something to be the case P. From this, they argue that P being the case would involve the absence of a reason or good explanation. Finally, from this it is inferred that P is in fact not the case. We can now see that IBE and PSR inferences are really just two sides of the same coin.

Here are some examples of PSR inferences in action. The following is from Spinoza’s proof of God or Nature in 1P11 of the *Ethics*:

Since, then, there can be, outside the divine nature, no reason, *or*, cause which takes away the divine essence, the reason will necessarily have to be in his nature itself, if indeed he does not exist. But it is absurd to affirm this of a Being absolutely infinite and supremely perfect. Therefore, there is no cause, *or* reason, either in God or outside God, which takes his existence away. And therefore, God necessarily exists, q.e.d. (91-2)

Here Spinoza argues that there is no reason or cause, neither from inside nor outside infinite substance, that is, a substance possessing all the attributes, that would keep it from existing, and so therefore infinite substance exists!

In arguing against Clarke and the Newtonian conception of absolute space, Leibniz also employs a PSR inference.

Now, from hence it follows (supposing space to be something in itself, besides the order of bodies among themselves) that it is impossible there should be a reason why God, preserving the same situations of bodies among themselves, should have placed them in space after one certain particular manner, and not otherwise—why everything was not placed the quite contrary way, for instance, by changing east into west. (325)

Here Leibniz argues that if space were an absolutely existing box with nothing outside it, then God would have no reason to arrange matter in it in one configuration versus that same configuration rotated 180 degrees.

### CONSEQUENCES FOR METAPHYSICS

I want to briefly describe some of the consequences of our acceptance of PSR inferences brought about by our acceptance of IBE. Here is Della Rocca again on contemporary metaphysics's rejection of the PSR and acceptance of brute facts:

What we find when we canvass contemporary philosophy is the prevalent presupposition that some facts are simply given and have no explanation. For example, in recent debates over the metaphysics of identity, it seems to be taken for granted on all or nearly all sides that there can be cases in which two things are distinct but in which their non-identity is primitive, without explanation. A commitment to primitive identity and non-identity is central to the otherwise widely divergent metaphysical views of David Lewis and Saul Kripke. (304)

An acceptance of the PSR and PSR inferences involves the rejection of primitive non-identity as a brute fact, that is, it involves an acceptance of the metaphysical Principle of the Identity of Indiscernibles—that if two things a qualitatively identical, then they are numerically identical.<sup>8</sup>

Spinoza endorses the Principle of the Identity of Indiscernibles. This is the principle—more often associated with Leibniz than with Spinoza—that if a and b are indiscernible, i.e., if a and b have all the same properties, then a is identical to b. One can see that this principle turns on the notion of explaining non-identity and, as such, one can see its roots in the PSR. Non-identities, by the PSR, require explanation, and the way to explain non-identity is to appeal to some difference in properties (47)

Finally, our acceptance of PSR inferences allows us to justify our acceptance of negative existentials. I claim not to believe in the Loch Ness Monster. If someone says to me, “Ahh, but you have no proof that the Loch Ness Monster does not exist!” I can say that I have all the proof I need in PSR inferences! I have no good reason to believe in the Loch Ness Monster, and this lack of explanation or reason is what justifies me in believing that it does not exist.

## CONCLUSION

I have argued for the notion of inferential presuppositions in order to explain Hume’s inductive circularity and Carroll’s deductive regress. I then argued that the inferential presupposition underlying IBE was the intelligibility principle, while the contrapositive of the intelligibility principle was the inferential presupposition behind the more controversial PSR inferences. This means that if we accept IBE we must also accept these more controversial PSR inferences. One of the things that makes PSR inferences controversial in the eyes of some, and useful in my opinion, is their use in establishing substantive metaphysical principles, like the Identity of Indiscernibles. An additional useful upshot in accepting PSR inferences is their use in justifying various beliefs involving negative existentials.

## NOTES

1. In the paper, the tortoise uses the example of the first proposition of Euclid, where he demonstrates the construction of an equilateral triangle: 1) Things which are equal to the same thing are equal to each other 2) The two sides of the triangle are equal to their common side 3) Therefore the two sides are equal to each other. The point is unchanged.

2. That is, within a system of natural deduction, certain rules will be basic and others derived. Often it might not make a difference which rules you include as basic. For instance, whether you include  $P \vee Q, \sim P \vDash Q$  or  $P \vee Q, P \supset R, Q \supset R \vDash R$  as the basic rule for **v-out** doesn’t ultimately matter to the system. Details will change depending on which you choose, but either will work. In choosing the former as basic the former will act as an inferential presupposition and vice versa. So whatever rule you end up choosing for **v-out** will be the inferential presupposition for the other rule, which is derived from it, and for itself, because it is a basic rule of inference. Thus modus ponens, as a basic rule, acts as its own inferential presupposition.

3. I want to set aside the question of whether IBE is a species of inductive inference or its own species of abduction. Hume has rightly shown us that induction works via causal mechanisms, so insofar as IBE does not do that, I think there is good reason to think it is its own unique form of inference. However, nothing I say here should depend on settling this issue now.

4. The distinct inferential presupposition here is yet another reason to think IBE is separate from induction in general.

5. Brief reflection will show that what I’m calling the intelligibility principle, that if  $P=T$  then there is a best explanation for P, is one version of what has been known in the

## Inference to the Best Explanation and the Principle of Sufficient Reason

history of philosophy as the Principle of Sufficient Reason (PSR)! Stronger versions would include the converse as well, that if there is a best explanation for P then P=T. See Spinoza's *Ethics* 1A3 and proof of God 1P11. Also Leibniz, "that nothing happens without a sufficient reason why it should be so rather than otherwise" (324).

6. There is no agreed upon list of features that make up our notion of best explanation. Usually included are features like simplicity and conservation, features that guarantee that we minimize modifications to our web of belief. In addition to obeying Bayes' Theorem in updating our web, our notion of bestness likely includes alethic and aesthetic considerations as well. That is, we want our best explanations to be both true and beautiful.

7. There are all kinds of things to say here. One might argue that the notion of an intelligibility of the world that does not match ours just means that the world is unintelligible. *À la Kant*, our IBE practices and the intelligibility principle that underlies them would be regulative ideals or "as if" pragmatic stances that we must take towards the world in order to have experience at all. I prefer to think of it more empirically, as us learning from the world and adjusting our notion accordingly.

8. The Principle of the Identity of Indiscernibles should not be confused with its converse, the *logical* Principle of the Indiscernibility of Identicals—that if two things are numerically identical, then they are qualitatively identical.

### WORKS CITED

- Aristotle. 1987. *A New Aristotle Reader*. Ed. by J. L. Ackrill. Princeton: Princeton UP.
- Carroll, Lewis. 1895. "What the Tortoise Said the Achilles." *Mind* 4 No. 14 (April): 278-280.
- Della Rocca. 2008. *Spinoza*. NY: Routledge.
- Hume, David. 1975. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. Ed. by P. H. Niddich. Oxford: Clarendon Press
- Leibniz, G. W. 1989. *Philosophical Essays*. Trans. by Roger Ariew and Daniel Garber. Indianapolis: Hackett.
- Spinoza, Benedict. 1994. *A Spinoza Reader: The Ethics and Other Works*. Ed. and Trans. by Edwin Curley. Princeton: Princeton UP.

