Transcript: Philosophy VIDEO 2.5 – Puzzles About Confirmation: Non-Black Non-Ravens

We're getting kind of big picture again, so I'd like to go small again. This video presents a puzzle, about the same size and shape of that card game I gave you before – you remember, the one where the solution was to disconfirm the rule?

This new puzzle I'm about to puzzle you with is intended as a response to the big picture stuff that came at the end of last video. Remember? I was considering, in an apocalyptic way, how this module—Q, Asking Questions—might fail. I concluded by saying even in the worst case it's still pretty interesting. Maybe we don't know how science works.

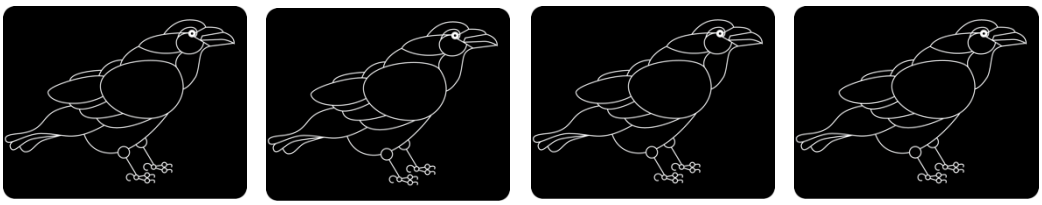
Now that result may seem just silly. Obviously we can do science. Being skeptical about the existence of science seems dumb. But note. I didn't say we would find out that we can't do science. I said maybe we would find out we don't know how science works. Not the same thing.

Remember: babies can ask questions. I'll bet they don't know what a question is. Maybe we are more like big babies than we know, where science is concerned.

Maybe we are doing—even doing great—but we know not what we do.

Now, the puzzle.

Let's start with good old Francis Bacon. Remember, he developed an inductive logic. It seems like we have to have some way of moving in a rational, warranted fashion, from particular observations to general conclusions. Sounds legit. But how does it go? How about this?

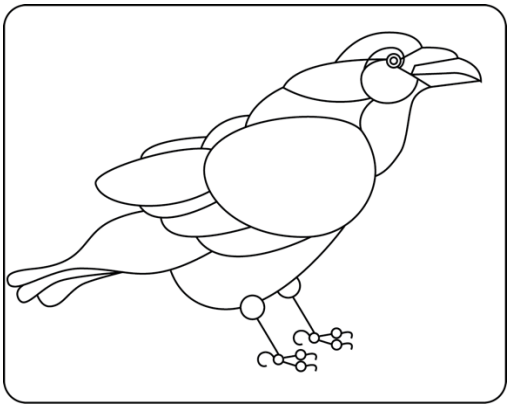


I see a black raven. I see another black raven. I see another black raven. I see another black raven. At a certain point along this line

All ravens are black

starts to sound pretty good.

The obvious problem with this is what we might call the albino raven problem. 1 in a million ravens is an albino raven, a white raven. How can you handle albino ravens, inductively?



But no one talks about albino ravens in this context. They use a different stock example, so let tell you what other example is:

black swans.

As you may know, unlike ravens, swans are white.

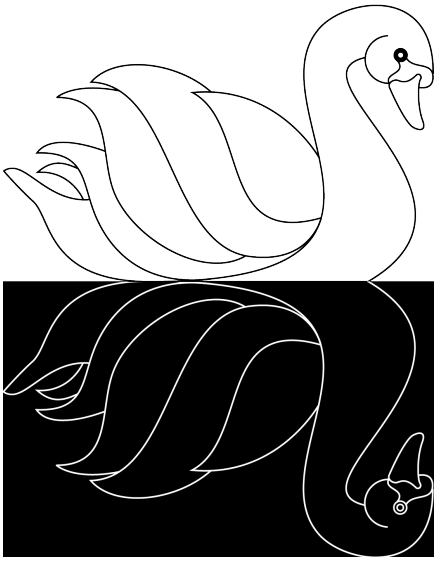
Well, European swans.

One white swan, two white swans, three white swans, four white swans. So long as you are in Europe,

All swans are white

is looking good. In fact, a famous Roman poet, Juvenal, used ‘black swan’ as a symbol for ‘thing that doesn’t exist’.

But Australian swans are black. Some of them anyway. But it took Europeans a while to get to Australia. 2000 years, roughly, after the poet Juvenal said there weren’t any, the first European spotted a black swan in Western Australia.



In honor of this, the term black swan has acquired not just cautionary force but semi-technical use. An economist—a finance guy, Nicholas Taleb—wrote a popular book, *The Black Swan*, in which he talked about black swan events.

The financial crisis of 2008, which I mentioned in the last video, is an example. Taleb hit the jackpot, publishing *The Black Swan* in 2007. Anyway, unlike literal black swans—which are elegant birds, but don't matter much to global finance—a black swan event, in Taleb's sense, has three features: 1) it's an outlier; 2) it's impactful; 3) looking back, we should have seen this one coming.

How can we better prepare ourselves, mentally, for black swans in this sense (as opposed to the literal bird sense)?

In Socratic terms, how can we know what we don't know? Especially when the thing we don't know has specific features.

Which thing, that you don't know, has these three features? That seems like an impossible question to answer.

It's like asking: which thing, that you have lost, is lost under your bed. If I knew even just that, it wouldn't be a thing I have lost anymore.

Taleb wants to model how to be more effective at being ignorant, given that you are going to stay effectively ignorant. That's a tough problem.

Let's model the sudden emergence of a black swan from the mix. Or a white raven. Same difference. I'll go with white.

Suppose I have a giant urn, containing 1 million balls. Big urn. I draw out 1 ball—black; another ball—black; another—black; another—black.

Maybe you start to think: this is a giant urn of 1 million black balls. I know, I know, it's premature, but you thought it, didn't you? If you pick 100, 1000 black balls, all in a row, you will grow even more confident.

So:

Hypothesis1 All the balls in this urn are black.

But consider a pair of counter-hypotheses.

Hypothesis2 All the balls in this urn, except for one, are black. There's one white ball.

Hypothesis3 All the balls near the top of this urn are black. But there's a layer of white balls at the very bottom.

I hope you see the way Hypothesis 2 is like my albino raven, and the way Hypothesis 3 is like black swans in Australia (just color-reversed). I hope you also immediately see the abstract form of the problem. 2 & 3 are inconsistent with 1. But merely drawing 1000 black balls from the top of the urn doesn't give you any more reason to believe 1 than 2 or 3.

Even so, seeing those black balls come up, one after the other—black, black, black, black, black—who WOULDN'T be surprised to see a white one after a couple hours of monochromic monotony? But why? Like I said: Hypothesis2 and Hypothesis3 are equally likely, in light of the evidence.

So the logic of induction is a puzzle. And now, having warmed you up, I'm going to give you a puzzle, concerning induction, that is like that card game we played before.

By the way, the author of this puzzle is Carl Hempel, who I mentioned before. One of the architects of the DN model—the deductive-nomological model of explanation in science. Which just goes to show that he didn't totally neglect induction. Without further ado, Hempel's Black Raven puzzle.

If Raven -> Black

That's the rule. It says, in effect,

All ravens are black.

It seems like, somehow, every black raven you see should incrementally increase your credence in this proposition. That's fancy talk for: make you believe it more. And rightly, it feels like. It seems it's rational, if you see a pattern in nature, to believe the pattern is real, hence will probably persist. It's not certain; it's a reasonable bet.

So every black raven confirms the rule. Except we need to add that confirmation is a matter of degree. In ordinary speech, we say 'that confirms it!' when we think we've got absolute lock. Here we are NOT talking like that.

Confirmation is like grains of sand making a heap. Enough confirmations make a heap of confidence. And, as is the way with heaps, every little bit helps. Every black raven, that is.

To repeat:

If Raven -> black

This rule is confirmed by every black raven we see.

But remember what I taught you about contrapositives? Every conditional is logically equivalent to its contrapositive. To construct the contrapositive, negate each side of the conditional, and flip their positions. So the contrapositive of If raven -> black is:

If non-black -> non-raven

All non-black things are non-ravens. Weird sentence, I know. Just think about it for a couple seconds. You will see, I think, that if all ravens are black (let's forget the albinos) then this weird sentence has to be also true. If ravens are black, then non-black things are not ravens.

If these two statements are logically equivalent—there is no possible world in which their truth values diverge, then anything that is evidence for the truth of one of them will be evidence for the truth of the other. Right?

So: if a black raven confirms the truth of

All ravens are black

it also confirms the truth of

All non-black things are non-ravens.

Right?

But consider—my shirt. It's some kind of purple. Let's call it purple. I'm totally sure it's neither black nor is it a raven. I'm wearing a non-black non-raven to cover my upper body. (No need to thank me. Modesty is my virtue.)

So my shirt confirms the proposition that all non-black things are non-ravens.

But if my shirt confirms that, then my shirt also confirms the thing that is logically equivalent.

Namely:

My shirt is evidence that all ravens are black.

But how could my shirt be a contribution to scientific ornithology? That makes no sense.

Someone looks at my shirt and says ‘Holbo’s shirt makes me somehow slightly more confident that all ravens are black.’ Where’s the sense in that? Well, I just told you where the sense is. The more serious question is: where’s the nonsense?

How does the logic of confirmation work, such that a black raven but not a purple shirt slightly confirms the proposition that ‘all ravens are black’?

Thousands of people have worried about Hempel’s non-black non-ravens. But there isn’t a simple solution. There isn’t a completely agreed-on explanation of why my shirt isn’t evidence that all ravens are black, if black ravens are evidence for that.

Going back to the start of this video: I said maybe scientists are like babies. They ask a lot of questions but don’t really know how questions work. That didn’t sound right but I’ve just given you evidence it’s true. Namely, there are lots of people over in the science faculty right now striving to confirm things—to provide evidence for general claims. But I doubt those people could provide you with a true, general account of how confirmation works. Generally. Weird, huh?

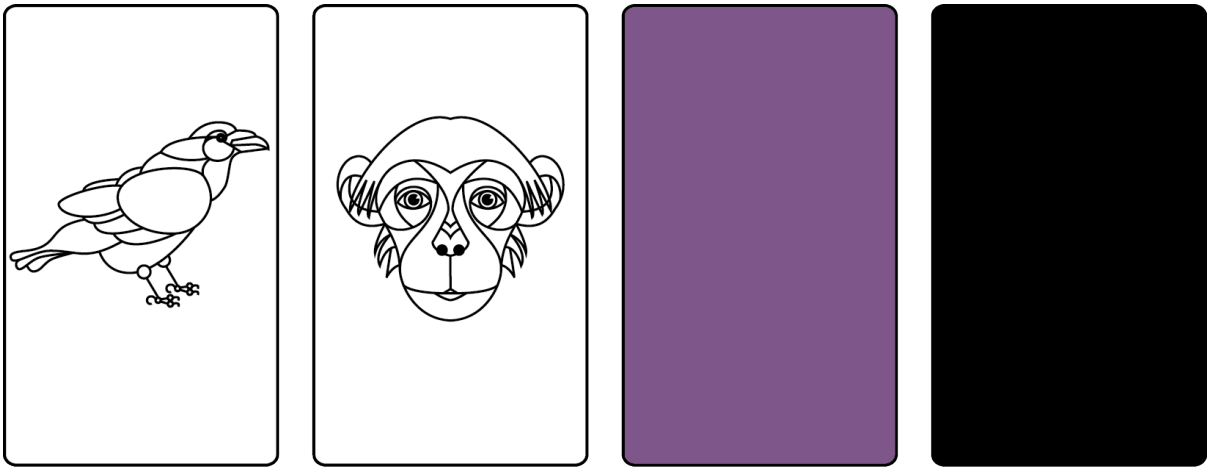
So let me try to explain something else. Is this non-black non-ravens thing exactly like that Wason Selection Task card game? How tight is the analogy? I think it’s pretty tight.

Suppose we’re playing cards again, and this time there are animals on one side, colors on the other.

Here’s the rule:

If raven -> black

Here are your cards:



We know our winning strategy: disconfirm.

So the cards we need to flip are raven—to see if there's something non-black on the other side; and purple—to see if there's a raven on the other side.

This result suggests there's something right about my purple shirt being important.

Here's a final clue. It's not a solution to the non-black non-ravens Hempel thing. Like I said, I don't have a solution. But I'm going to pull on this Wason Selection Task thread—this hint that maybe focusing on purple things is not as crazy as it sounds.

One thing that makes life not like cards is that you don't normally see the color of something, THEN what kind of thing it is, by flipping it. Or what kind of thing it is, THEN the color. You usually get both all at once or neither. But let's imagine a setting—non-card setting—in which they come apart.

Imagine you and your crack team of ornithologists have traveled to Raven Island—so-called because it's covered with ravens. You are tracking through the savage jungles of Raven Island, looking for unusual specimens. Anything new you can write home about. What you see is a whole lot of black, all around. Black black black black black black black.

Suddenly, through the trees you see a flash of purple. Eureka! Could it be? A purple raven! You'll be famous! You are off like a shot, racing after this rare specimen, possible Latin names running through your head. *Corvus corax purpura*?

When you catch it you realize it was just the team philosopher, Holbo, that idiot, who had wandered off and gotten lost. You'd glimpsed his purple shirt through the trees. That stupid purple shirt he always wears. Rats. I guess ravens are all black, after all.

I think this brings out how in a weird sense, a non-black non-raven—like my purple shirt—could actually confirm that all ravens are black. But the logic of confirmation is still puzzling. In the next video I'm going to suggest that maybe the problem is that we're not sure what question we are asking. Wouldn't that be a thing?

And one last, last note. I came up with this purple shirt example totally on my own, for use in this video. Then I found a paper in which a psychologist named Raymond S. Nickerson basically said the exact same thing, plus with extra math. The paper is "Hempel's Paradox and Wason's Selection Task: Logical and Psychological Puzzles of Confirmation", published in *Thinking And Reasoning* 1996. Kids, you should always give credit when someone else came up with it first. It's the right thing to do.

Correction: Holbo just cited a paper by Nickerson but he actually meant to cite a different paper, "Hempel Meets Wason", by I.L. Humberstone, from *Erkenntnis* 41, 1994. They are both good papers, so no harm done; but Holbo regrets the error. (Holbo gets confused sometimes.)