Inductive Knowledge

Andrew Bacon*

July 26, 2018

Abstract

This paper formulates some paradoxes of inductive knowledge. Two responses in particular are explored: According to the first sort of theory, one is able to know in advance that certain observations will not be made unless a law exists. According to the other, this sort of knowledge is not available until after the observations have been made. Certain natural assumptions, such as the idea that the observations are just as informative as each other, the idea that they are independent, and that they increase your knowledge monotonically (among others) are given precise formulations. Some surprising consequences of these assumptions are drawn, and their ramifications for the two theories examined. Finally, a simple model of inductive knowledge is offered, and independently derived from other principles concerning the interaction of knowledge and counterfactuals.

Philosophers have long been interested in the probabilistic structure of inductive inference — how the discovery that a great number of Fs are Gs can make it rational to be confident that all Fs are Gs. But there is another aspect of inductive inference that is not so often discussed — namely, that after making enough such discoveries we are eventually in a position to *know* that all Fs are Gs.[1] For I take it that we know that all emeralds are green (say) and, moreover, that this fact was first discovered on the basis of a sufficient number of observations of green (and only green) emeralds. Although the probabilistic and epistemic inferences are related, a theory of one sort of inference does not obviously provide a theory of the other.

One difference is that Bayesians typically assume that whether it is rational to be confident that all emeralds are green on the basis of your observations does not depend on whether it is in fact a law that emeralds are green. Whether

*Many thanks to John Hawthorne for some very helpful feedback on an earlier draft of this paper. I would also like to thank Jeremy Goodman, an anonymous referee for this journal and an audience at Princeton at which a version of this paper was presented.

[1]Here and throughout I shall use the expression 'S is in a position to know that P' to mean, roughly, that there is a way for S to form a belief that P (for the right sorts of reasons, etc.) that would constitute knowledge. Beyond this gloss, I will not attempt to give any sort of further analysis of this notion.

you are in a position to know, on the other hand, presumably does depend on whether it's a law.[2] Consider the following case:

> All Green by Fluke: It is a law that emeralds are either blue or green, but the distribution of colors is otherwise determined randomly. By chance it happens that all the emeralds in the actual world are green.

Even if all the emeralds are green, and I observe 100 green emeralds in a row, I take it that I am not in a position to infer that the next emerald will be green. In All Green by Fluke (henceforth, simply Fluke) it could easily have been the case that the next emerald is blue, and so a belief that the next emerald is green, although true, is not *safe*.[3]

Here's a natural (albeit somewhat idealized) model of Fluke inspired by safety accounts of knowledge: epistemic possibilities are identified with assignments of colours to emeralds, and an epistemic possibility w is consistent with my knowledge if at most n different emeralds in that possibility have colours that differ from their actual colours (n is a fixed number, which we shall call the *margin of error*).[4] After observing the first 100 emeralds to be green, and ruling out sequences that don't begin with 100 green emeralds, there are still worlds that are open where some of the remaining emeralds — up to n of them — are blue. This sort of model correctly predicts that I am not in a position to know that all emeralds are green after learning that the first 100 are green.

Here is a second sort of case which, I take it, is more conducive to inductive knowledge:

> All Green by Law: It is a law that all emeralds are green.

Let us suppose that at the beginning of investigation both the hypothesis that emeralds have colors randomly and the hypothesis that they're all green by law (henceforth, Law) are open possibilities. In particular, at the start of the investigation it's true, for all you know, that some of the emeralds are blue. [5]

If knowledge by induction is ever possible it seems like it should be possible in this case: there is some critical N such that after observing that the first

[2] Of course, if your evidence just is your knowledge (as in Williamson (2000)), and rational confidence is guided by the degree to which it is supported by the evidence, then there is less distance between the questions. In that case, the question concerning how we get to acquire inductive knowledge/evidence is even more urgent.

[3] In what follows I assume that there's a difference between a world where all emeralds in the world are green by law and a world where they're all green by fluke. This contradicts a crude version of Humeanism about laws in which the law that all the emeralds are green, if it is a law, holds purely in virtue of the fact that all emeralds are green. However nothing in this paper essentially turns on this rejection of Humeanism: instead of considering the world in which all emeralds are green by fluke — a world the crude Humean rejects — we can , for most purposes, just focus on worlds in which the first 100 emeralds are green by fluke, of which the Humean in question accepts many.

[4] I defend this model of these sorts of cases in Bacon (2014).

[5] One epistemic possibility in which there are blue emeralds is one in which it's a law that all emeralds are blue. But if the only possibilities were ones in which it was a law either way, we would be able to conclude the colors of all the emeralds from the color of the first observed emerald. Thus a plausible model of our ignorance about inductive knowledge had better include some worlds where the emeralds have different colors.

N emeralds are green, we can knowledgeably conclude that all emeralds are green. Because it's a law that all emeralds are green, our true belief that all the emeralds are green appears to be safe in the relevant respects. In worlds where LAW is true there is some N such that:[6]

Inductive Knowledge After observing emeralds $e_1 ... e_N$ to be green we are in a position to know that it is a law that all emeralds are green.

In what follows we will be concerned with trying to explain Inductive Knowledge.

The possibility of inductive knowledge, however, is undermined by a certain sort of skeptical argument. The argument rests on two popular ideas. The first is an intuition about 'lottery-esque' scenarios like FLUKE: if the emeralds colors are distributed at random, then prior to observation I'm not in a position to know that any particular sequence of colors won't be observed. In particular, I'm not in a position to know that the emeralds won't all be green by fluke. This intuition is clearly of a piece with the popular idea that we can't know that a particular ticket will not win a fair lottery, even when the number of tickets is very large. If this thought is good in FLUKE, it seems it is also good in a situation where you don't know whether the emeralds' colours are distributed at random, or according to a law: if, for all you know, the colors are distributed at random, then it seems we're not in a position to know that any particular random distribution won't occur. Thus even if LAW is in fact true, we should maintain that:

Lottery Intuition Prior to making observations, I'm not in a position to know that the emeralds won't all be green by fluke.

The second ingredient for our skeptical puzzle follows from the following natural idea: that if P is true for all you know, and is consistent with some evidence E, then (ordinarily) P ought to be true for all you know after coming to learn E and nothing else.[7] In particular:

(*) If I'm not in a position to know that the emeralds won't all be green by fluke, I'm not in a position to know this after observing that emeralds $e_1 ... e_N$ are in fact green.

After all, the evidence that $e_1 ... e_N$ are green is entirely consistent with the open possibility that all emeralds are green by fluke. Indeed, (*) follows from a more specific account of what we learn when we observe an emerald along with a closure assumption. Since this principle will play a role in the rest of the paper lets introduce it now:

[6] Of course, there are some who might be tempted to a sort of skepticism about inductive knowledge. Indeed, some philosophers have raised skeptical challenges for our knowledge about the future more generally, for non-deterministic worlds that are governed by chance (see, for example, Hawthorne and Lasonen-Aarnio (2009)). The issue here is more pronounced, however, since we are entertaining the idea that even when when something is in fact governed by a law, the mere possibility of a non-lawlike world, such as an ALL GREEN BY FLUKE world, is sufficient to undermine knowledge about the future from induction. Despite raising the challenge, Hawthorne and Lasonen-Aarnio see skepticism about the future as a last resort.

[7] This general principle must be qualified: see section 3 for a more comprehensive discussion.

What You See is What You Know When you observe an emerald, you get to know that colour it is and nothing more. Your knowledge after observing the emerald is just whatever your knowledge was before observation, plus the proposition that that emerald is green.

It is illustrative to see why (*) follows from this principle with closure. In sketch, suppose I'm not in a position to know that the emeralds won't all be green by fluke. Given multipremise closure this means that the FLUKE possibility is compatible with the conjunction of what I'm in a position to know. By What You See is What You Know and closure, after observing emeralds $e_1...e_N$, I'll be in a position to know the conjunction of my prior knowledge with the claim that $e_1...e_N$ are all green, which must also be compatible with the FLUKE possibility. (A less sketchy version of this argument appears in section 4.2.)

This gives rise to our first skeptical 'paradox':

Skeptical Puzzle 1 The Lottery Intuition and (*) are inconsistent with Inductive Knowledge.

Lottery Intuition is the antecedent of (*), so by modus ponens they entail that I'm not in a position to know that all emeralds are green after observing $e_1...e_N$ to be green. Thus Lottery Intuition and (*) are inconsistent with Inductive Knowledge, so, as a matter of logic, an anti-skeptic must deny one of these two intuitions.

This effectively leaves us with two anti-skeptical options consistent with closure.[8] One is to reject the Lottery Intuition, and maintain that we are in a position to rule out certain possibilities, like FLUKE, prior to making any observations. Another is reject (*): some observations confer more knowledge than that the observed emerald is green.[9] Let's give these options names:

> **Prior Knowledge**: Prior to observation you are in a position to know that some sequences of colours don't obtain if emeralds have colours at random.
>
> **Bonus Knowledge**: Some (possibly all) observations confer more knowledge than the proposition that the observed emerald is green.

We will consider both sorts of theories in what follows, but it is important to bear in mind that the dilemma between rejecting the Lottery Intuition and (*) is forced on us once we have accepted the possibility of inductive knowledge. The rest of the paper will be devoted to further issues pertaining to these two responses to the paradox.

[8] Rejecting closure opens up other ways to respond to the paradoxes, but in this paper we shall limit our discussion to views that accept closure.

[9] Note that anyone who keeps the Lottery Intuition and Inductive Knowledge must reject (*), whether they accept closure or not. Thus all theories who accept the Lottery Intuition must accept some version of the bonus knowledge theory below: prior to observation it's true for all you know that $e_1...e_N$ are green by fluke, but after observing $e_1...e_N$ to be green you have the extra piece of knowledge that it wasn't a fluke.

In section 1 I impose some assumptions. In section 2, I outline two theories concerning how inductive knowledge is possible, and show that any theory accepting Inductive Knowledge must conform, broadly speaking, to one of those two sorts. In section 3 and 4, I outline some puzzles for both theories. In section 5, I highlight some inferences connecting counterfactuals and knowledge and in section 6 apply them to some of these puzzles. In section 7, I leverage the relationship between counterfactuals and knowledge to explore a specific version of the second sort of theory introduced in section 2.

Throughout this paper I will introduce various principles, like What You See is What You Know, that have some plausibility in the particular cases at hand. It would be wrong-headed, however, to read such principles as exceptionless generalizations. The challenge is to explain how they could fail, if they indeed fail, in these cases.

1 What is Inductive Knowledge?

Let me begin by emphasizing that not all cases of acquiring knowledge of laws will count as inductive by my standards. At one extreme, I imagine some knowledge of laws could in principle be made from observing a single instance of a law. For example, if I had been told by a trustworthy person prior to observing any emeralds that either it was a law that emeralds are blue or a law that they're green, I could know which by making a single observation. Indeed, perhaps we are born with a natural sensitivity to what sorts of properties are likely to be subject to laws. Perhaps I have innate knowledge that whether animals lay eggs or give birth to live young is likely to be subject to a law one way or the other. In that case, I might knowledgeably infer that it's a law that ducks lay eggs after observing only one duck lay an egg (since I then know which law is in operation). The sorts of cases of inductive knowledge I am interested in are cases where, prior to observation, I do not know whether there's going to be a law of some sort, or no law of that sort — in other words, cases where knowledge from one case is not in general possible.[10]

I take it that these sorts of cases of inductive knowledge are pervasive, and representative of the way that science usually proceeds. To illustrate the sort of structure we are interested in, consider the following well-known example from Mendelian genetics. Mendel investigated two types of pea plants, the F_1 generation and the F_2 generation pea plants[11], and discovered that F_1 generation pea plants always had purple flowers, whereas the F_2 generation pea plants either had white flowers or purple flowers at random (with a ratio of 3 to 1 purple to white on average).

[10] On these grounds, one may take exception to my choice of example: perhaps the colors of precious stones are the sorts of things we are typically in a position to know will be subject to laws. I entrust it to my reader to see how the remarks that follow can be applied to whatever example of inductive knowledge they favor.

[11] F_1 generation pea plants are obtained by interbreeding a (purebred) white flower pea plant with a purple flower pea plant. F_2 generation pea plants are obtained by self-fertilizing a F_1 generation pea plant

On the one hand, it seems Mendel should have been in a position to know that it was a law that the F_1 generation plants are always purple flowered after performing a sufficient number of experiments. On the other hand, it's not plausible that he was in a position to know, prior to investigation, that there would be some color such that it was a law that the F_1 generation plants would that color, since it is not a law that the F_2 generation plants are any particular colors, and he had no way of knowing that the F_1 generation wouldn't be like the F_2 generation in that respect. Thus, prior to investigation there were at least two sorts of possibilities open to Mendel: possibilities where it is a law that the F_1 generation plants are a certain colour (white or purple), and possibilities where the F_1 generation plants are white or purple flowered at random,[12] and that after observing enough purple flowered F_1 generation plants he was in a position to rule out the worlds where it was random.

In order to productively investigate the structure of inductive knowledge we shall impose some simplifying assumptions. Firstly, we shall focus on special cases of inductive knowledge having the structure of the example from Mendelian genetics discussed above. Although we will tailor our discussion to this particular sort of example, the case has an underlying structure that is common to many instances of inductive knowledge. Thus there is also a reasonable expectation that a detailed investigation of this idealized sort of case will yield interesting conclusions about the structure of inductive knowledge more generally.[13]

Secondly, we shall make some fairly strong idealizations about the logical competence of the agents we shall consider. In the present context, these idealizations are quite reasonable: we are, after all, interested in a form of *empirical* knowledge acquisition and so it is natural to focus on the case where the agent doesn't lack any relevant *a priori* knowledge. We shall often also make the assumption that knowledge is transmitted across single and multi-premise inferences — an assumption that is contentious even when applied to logically flawless agents.[14] In the context of induction, probabilistic accounts of knowledge naturally come to mind, according to which a proposition is known if it is true and its epistemic probability exceeds some threshold. Such accounts provide principled reasons for giving up multipremise closure, but are not nec-

[12]The latter possibilities include worlds where there are still laws about the proportion of certain colors to others.

[13]That said, I think the anatomy of inductive knowledge of this particular is instructive in its own right.

[14]According to some philosophers (see, e.g., Nozick (1981)) knowledge fails to be preserved by competent deduction, so that even a logically omniscient agent — someone who correctly infers and comes to believe all the relevant logical consequences of their knowledge — may still fail to know all the logical consequences of their knowledge. Note, finally, that even assuming a logically flawless deducer, and granting that knowledge is usually transmitted across competent deduction, one might still want to qualify closure. For example, suppose I believe both P and Q, but only know P. If I can deduce R from both P and Q, and by luck I end up inferring it from P, you might think my belief in R doesn't count as knowledge, since I wouldn't have known it had I inferred it from Q. These sorts of exceptions will not play an important role in the following discussion.

essarily plausible as theories of knowledge.[15] While I in fact subscribe to a suitably qualified version of single and multi-premise closure for knowledge, I will not attempt a serious defense of that assumption here: one may regard this paper rather as an exploration of the options once closure is held fixed.

The circumscribed case of inductive knowledge that will be our focus can be described as follows. In the general setting, we assume that there is a number of distinct objects, $a_1, a_2, ...a_n$, belonging to a certain kind, F. We also assume that there is a set, S, of pairwise incompatible, natural properties (we are ruling out properties like *being grue*) and that each F has exactly one property in S. To represent the agent's ignorance about the laws, and about whether there are any laws, we must accommodate at least two types of world. In *lawlike* worlds, all the Fs possess a single member of S, and it is moreover a law that all Fs possess that member of S. In what I'll call *random* worlds, there are no laws of this sort, and different Fs may be observed to have different members of S. I do not assume that randomness is analysed in terms of objective chance: the S-properties are assigned 'randomly' only in a sense that it is not a law that they are assigned in that particular way. Crucial to our discussion is the fact that there can be worlds where all Fs do have the same natural property but it is not a law that they do. In these cases we will say that the Fs all have that property G *by fluke*. Again, we are not assuming a prior notion of 'fluke': this simply means that the fact that all Fs are Gs is not entailed by a law that holds at that world.

This assumption looks like it might be in conflict with the crude form of Humeanism mentioned in footnote 3. Note, however, that to make the original modeling assumption is not necessarily to prejudge Humeanism as a metaphysical thesis: the worlds in our model are only supposed to represent epistemic possibilities and the existence of a world in which certain patterns hold without the corresponding laws does not necessarily mean that that world represents a genuine metaphysical possibility. Moreover, even the Humean can accept that it might fail to be a law that all Fs are G even if all Fs that could ever be observed are G by fluke. So we can actually give our model a Humean-friendly interpretation even if we insist that the worlds are metaphysical possibilities: on this interpretation $a_1...a_n$ correspond to the emeralds that could eventually be observed by us (e.g. the emeralds in our solar system), random worlds are worlds where there's at least one green and at least one blue emerald (possibly unobservable) and the lawlike worlds are the worlds where all emeralds, observable and unobservable, are a single color.

For the sake of concreteness, let's work with a more specific example: let us suppose that we are observing emeralds and that there are only two possible colors an emerald can have — green or blue. (In this respect the case is exactly parallel to the pea plants, since one could plausibly have known prior to the experiment that the F_1 generation pea plants would be either white or purple

[15] Or so I argue in Bacon (2014) §5, in a similar context. A more modest version of the probabilistic account merely asserts that high probability is necessary for knowledge. Note, however, that the mere necessity claim is silent about multipremise closure and is consistent with its truth and its falsity.

flowered.) Suppose also that we know that there are exactly one thousand emeralds, and that we know which emeralds they are. If one liked one could imagine this stipulation as the result of God having told us these facts prior to gaining any empirical knowledge. The resulting considerations then pertain to the effect that further observation of emeralds would have on this knowledge. We shall write $e_1...e_{1000}$ to denote the 1000 emeralds, and for each $i \in \{1...1000\}$ we shall write G_i to denote the proposition that e_i is green.

2 Two Theories of Inductive Knowledge

The upshot of Puzzle 1 was that there are, broadly speaking, only two sorts of views that validate Inductive Knowledge: the Prior Knowledge Theory (that gives up the Lottery Intuition) and the Bonus Knowledge Theory (that gives up (*)). The first theory maintains that, in worlds in which it's a law that all emeralds are green:

Prior Knowledge Prior to observing any emeralds I am in a position to know the falsity of All Green by Fluke: I'm in a position to know that it's not the case that emeralds $e_1...e_N$ are all green unless it's a law that they're all green.

In other words, one is in a position to know, prior to investigation, that if emeralds $e_1...e_N$ were assigned colors randomly, they wouldn't all end up colored green by fluke. The intuition here is similar to the intuition that if a coin is to be flipped a thousand times, we are usually in a position to know that the coin won't land heads every time.[16] Of course there are some very rare exceptions: if the coin does land heads every time, or lands heads almost every time, then the belief is either false or could easily have been false, and so in these cases such a belief presumably won't count as knowledge. But in statistically normal cases this intuition seems to be in good standing.

The prior knowledge theory states that in cases where inductive knowledge is possible we are in a position to know that the emeralds $e_1...e_N$ aren't all green by fluke before making any observations. It does not say that this is *a priori* knowledge, since there may be some worlds where we aren't in a position to have this knowledge prior to investigation. Much like in the coin example, in a world where the emeralds are all green by fluke, or nearly all the emeralds are green by fluke, our belief that it's not the case that all the emeralds are green unless they're all green by law is either false or could easily have been false, and so we are not in a position to know this.

The prior knowledge theory has a straightforward explanation of Inductive Knowledge. Upon observing that the emeralds $e_1..e_N$ are green, I get to know

[16] Although note that this intuition, if it is indeed an intuition, is in tension with other intuitions sometimes raised in the context of the lottery paradox, namely: that we are not in a position to know that any particular ticket will lose in a fair lottery. One might think that a coin landing heads a thousand times is relevantly analogous to winning a lottery (with a very large number of tickets). I defend the view that we can know lottery propositions like this in Bacon (2014) by appealing, among other things, to the possibility of inductive knowledge.

that $e_1 ... e_N$ are all green. Given closure of knowledge under competent deduction, I can thus knowledgeably conclude, with the help of my prior knowledge, that they are not all green by fluke. In other words: we get to know that it is a law that they are all green. (Recall that we are using the phrase '$e_1 ... e_n$ are green by fluke' to simply mean that $e_1 ... e_n$ are all green, but the laws do not entail this.) It follows that the prior knowledge theory can uphold the simple theory about the effect of observation on knowledge which we introduced in the introduction:

What You See is What You Know When I observe that the ith emerald is green I come to know that the ith emerald is green and nothing more: my new knowledge is just the conjunction of my old knowledge with the proposition that the ith emerald is green.[17]

This principle may be recast in a simple possible worlds framework for modeling knowledge as follows: If my knowledge before observation is represented by a set of worlds, K — the worlds that obtain for all I know — then my knowledge after observing that the ith emerald is green is $K \cap G_i$ — the set of worlds in K in which the ith emerald is green. This possible worlds model will not be assumed in what follows, but it is helpful for fixing ideas.

What You See is What You Know is not supposed to be a general theory about the effects of observation on knowledge. One could imagine a sort of reverse of our setup: that I begin with knowledge that we are in a random world, and then by coincidence observe 100 green emeralds in a row, and come to know that they are green. One might think that at some point my new knowledge would undermine my initial knowledge that the emeralds have random colors.[18] In that case, my knowledge would fail to always increase in the way that What You See is What You Know predicts. Nonetheless, I submit that What You See is What You Know provides an extremely compelling account of the effects of observation on our knowledge in the case at hand.

It is quite hard to avoid the prior knowledge theory without giving up Inductive Knowledge. If All Green by Fluke were possible prior to investigation — i.e. if it was possible that emeralds had their colors randomly and were all green by happenstance — then presumably it would still be a possibility after learning that the first N emeralds are in fact green. For the new evidence — that emeralds $e_1 ... e_N$ are green — is completely consistent with the possibility that they are all green by fluke, so if that was a possibility beforehand the new evidence has done nothing to rule it out. But it then follows that we are not in a position to know that it's a law that emeralds are green after the first N observations after all.

In the introduction we argued that this followed from closure and What

[17] As currently stated What You See is What You Know simply presupposes closure. The anti-closure theorist, by contrast, may model my knowledge at a time by a set of propositions. What You See is What You Know may be reformulated in that setting as the idea that my knowledge after observation is just the result of adding the proposition that the ith emerald is green to the set of propositions I knew before I made the observation.

[18] Although even this is contentious; see Lasonen-Aarnio (2010).

You See is What You Know. To illustrate that argument, consider, again, the simple possible worlds model of What You See is What You Know. If my initial knowledge is represented by a set of worlds K, then by What You See is What You Know my knowledge after making the observations is $K \cap G_1 \cap ... \cap G_N$ — i.e. the set of worlds belonging to K where $e_1...e_N$ are all green. But if a world belongs to K and to G_i for each i from 1 to N, then it clearly belongs to $K \cap G_1 \cap ... \cap G_N$. In particular, if a world where $e_1...e_N$ are green by fluke belongs to K then it must belong to $K \cap G_1 \cap ... \cap G_N$, my knowledge after making the observations.

However there is an alternative theory, which we've dubbed the *bonus knowledge theory*, that rejects prior knowledge (and consequently What You See is What You Know or closure). According to the simplest version of this theory, prior to investigation every consistent distribution of colors of emeralds and laws is possible, including FLUKE — the possibility that the observed emeralds are all green by fluke. However observing that a certain number of emeralds are green, on this view, will suffice for one to be able to rule out FLUKE despite the fact that the evidence seemingly conferred by those observations is completely consistent with it.[19]

Bonus Knowledge Prior to investigation it is epistemically possible that all N observed emeralds are green by fluke. One or more of the observations confers more knowledge than that the observed emerald is green.

Indeed, it follows that there is some N such that after observing the Nth green emerald one is in a position to rule out the possibility that the N observed emeralds are green by fluke, even though this was not known before actually performing the observation. This is, admittedly, fairly counter-intuitive. For intuitively, all one learned upon making the Nth observation was nothing more than that the Nth emerald is green, and this newly acquired evidence is completely consistent with the possibility that all N emeralds were green by fluke. If this was a possibility before the observation, it seems, the new evidence cannot be used to rule it out.

One way to make sense of this is by maintaining that a particular observation N provides one with a stronger kind of evidence than the earlier observations. While all I learn from the ith observation, when $i \neq N$, is that the ith emerald is green, the Nth observation provides me with more knowledge: that the Nth emerald is green *and* that these emeralds weren't all green by fluke.

This model is extremely unattractive, since it breaks an apparent symmetry between the observations: it violates the intuition that the observations are all equally informative. This suggests that a more natural model would be one in which each observation directly confers the same amount of extra knowledge, and that these extra bits of knowledge, when conjoined, entail (possibly along with some prior knowledge) that the emeralds are not green by fluke. For example, after observing the ith emerald it's plausible that one learns both that the ith emerald is green *and* that it has been observed by you. Perhaps you're

[19] Here and throughout I use 'rules out that P' to just mean 'knows that P is not the case'.

not in a position to know that some particular emeralds, $e_1...e_N$, won't all be green by fluke prior to observation, but perhaps you do know that it would be too much of a coincidence for the emeralds that you will observe to be all green by fluke. (More generally, theories of this sort may posit that there is some property, F, such that observing an emerald puts one in a position to know that the emerald is green and F, and moreover one knows prior to investigation that the emeralds wont all be green and F by fluke.)

It's interesting to note that this theory does not differ substantially from the prior knowledge theory when applied to slight variants of the set up. One could imagine a variant of the present thought experiment in which the observer knew in advance which particular emeralds were going to be observed (but not their colours). Armed with this extra haecceitistic information about which particular emeralds are to be observed, the impact of each observation is to add to your knowledge only the proposition that the emerald e_i is green, since the proposition that e_i is observed by you is already known. In the variant experiment, then, the observer knows prior to observation that $e_1...e_N$ won't be green by fluke, just as the prior knowledge theory maintains.

The question of whether this is a version of the prior knowledge or bonus knowledge theory is not a productive question on its own, since it depends on how we set the observations up. But given the similarities to the prior knowledge theory it's natural to ask whether there is anything to recommend it over the regular version of the prior knowledge theory outlined above.

It is worth reminding ourselves that there is no causal connection between which emeralds are observed and whether or not they are green. It is therefore extremely surprising to think that, even though it's open at the start of the investigation that $e_1...e_N$ are all green by chance, we can nonetheless know from the start that these emeralds won't all be green by chance if we observe them. How could looking at them make a difference? This point is more dramatic when we consider the observation of e_{N-1}: the observation just before the observation that puts us in a position to know that it's a law that emeralds are green. At this point it's true for all you know that it's not a law that emeralds are green but that e_N is green by chance, but you also know that it won't be green by chance if you look at it. Since this sort of phenomenon is not predicted by the regular prior knowledge theory, this seems like a strong reason in favour of it.

In light of this, let's consider a final version of the bonus knowledge theory. According to this version, the observations all directly provide one with the same amount of non-inferential knowledge: perhaps that the ith emerald is green (the proposition G_i). The propositions $G_1...G_N$ then jointly license the inference to the conclusion that it's a law that all emeralds are green, from which one can *inferentially* acquire the knowledge that all emeralds are green. But one must bear in mind that this is cannot be a logical inference since $G_1...G_N$ do not entail this conclusion even (according to this theory) in conjunction with your prior knowledge. Moreover, it is presumably not an inference that can be legitimately made from any proper subset of $G_1...G_N$, assuming N is the smallest number of observations needed to provide us with inductive knowledge. This picture upholds the symmetry intuition since, by changing which observation was made

last, the critical observation can be shuffled so that it occurs at the observation that e_i is green, for any emerald e_i. It also keeps on to some version of the idea that all the observations are equally informative: each observation provides you with the same non-inferential knowledge — that the ith emerald is green — even though there will be a particular observation that tips the balance and allows you to acquire, inferentially, the knowledge that emeralds are green as a matter of law.[20]

We shall mainly focus on versions of the bonus knowledge theory in which all sequences of colours are open prior to investigation. One could, in fact, consider an intermediate version of the bonus knowledge theory in which some sequences of colors are ruled out beforehand, just not sequences in which the first 100 emeralds are green by fluke. This sort of view, however, doesn't have much to recommend itself: it accepts some of the unintuitive features of the prior knowledge theory, in addition to the unintuitive features of the bonus knowledge theory, but does not put them to any sort of use.

3 Puzzles for the Bonus Knowledge Theory

Both the positions discussed above bear at least a formal analogy with a debate in the epistemology of perception. According to the *dogmatist* it is epistemically possible, prior to any observations, that I be a handless brain in a vat and that I have the perceptual experience of having hands. But according to these philosophers, when I actually learn that I'm having the perceptual experience of having hands I also get to know that I have hands, and thus get to know that I'm not a brain in a vat.[21] Just as in the inductive case, the evidence — that I'm having the perceptual experience of having hands — is consistent, given my prior knowledge, with the hypothesis that I'm a handless brain in a vat, but upon receiving this evidence I'm in a position to know the skeptical scenario doesn't obtain. In this section I'll draw on some of the critical literature on dogmatism, and show that a similar set of issues arise for the bonus knowledge theory.

First, consider the following natural principle about knowledge:

> If prior to receiving the evidence E it's true for all you know that E and not-P, then you cannot come to know P by learning E and inferring it from E and things you previously knew.[22]

[20] This list of versions of the bonus knowledge theory is not exhaustive. There is also a picture that doesn't draw on the distinction between inferential and non-inferential knowledge: you are just in a position to know certain things at certain times, in the sense that, if you were to believe them those beliefs would constitute knowledge (whether you believed them on the basis of an inference or not). On this picture, after some number of observations, you are just in a position to know that it's a law that emeralds are green. I won't attempt to exhaustively characterize the different versions the theory; most of what I say is not sensitive to which version we are concerned with.

[21] See Pryor (2000) and Huemer (2001).

[22] I have taken this principle from Dorr et al. (2014) §6. As they note, it also needs to be qualified in various respects to be remotely plausible. For example, if prior to receiving

This principle is implicit in What You See is What You Know. However, since What You See is What You Know is a theory of knowledge acquisition the bonus knowledge theory explicitly rejects, it does not follow that the bonus knowledge theory should conform to this principle. Indeed, this principle is not consistent with the inferential version of the bonus knowledge theory discussed in the last section: before I receive the evidence that the emeralds $e_1...e_N$ are green it's epistemically possible that the emeralds $e_1...e_N$ are green but by fluke, yet after learning that $e_1...e_N$ are green, I knowledgeably infer that they're green by law not by fluke.

To dramatize why this result might be unwelcome, we might make a loose comparison to a certain sort of Dutch-book argument used to show that if you know your credences will end up a certain way, they should already be that way.[23] Suppose that it is in fact a law that emeralds are green, and suppose that a bookie offers you a bet that costs \$2 and pays out some sufficiently large sum if the first N emeralds are observed to be green, but it was by fluke (e.g. there are some blue emeralds you could have observed but didn't). Here N is some number known to be greater than or equal to the critical number of observations needed to know that there's a law that emeralds are green.[24] Call this critical number the *tipping point*. Since, for all you know to begin with, the first N emeralds are green by fluke, and the sum is sufficiently large, you should buy the bet. However, the bookie knows that anyone who has observed N green emeralds will be in a position to know that they were green by law not fluke (since N is known to be at least as big as the tipping point). She also knows you'll sell the bet back to her if you know it to be worthless. If the bookie strategizes so as to offer to buy the bet back off you for \$1 if the first 100 emeralds turn out to be green, she can know that she'll make money off you no matter what. For if the first 100 emeralds are not all green, she has made \$2, and if they are, she knows you will sell the ticket back for \$1 (for she knows you will know it to be worthless after observing that many green emeralds), leaving her with a profit of \$1. She will make a profit of 1 or 2 dollars either way, and she can know this.

The principle has a certain amount of pretheoretic appeal, although it is not out of the ordinary for some piece of common sense to be overturned in light of these sorts of considerations. However, the bonus knowledge theory requires an even greater departure from common sense than the above suggests. As essentially pointed out by White (2006) in the context of the dogmatism, it also conflicts with the weaker principle:

> If you are not in a position to rule out P before receiving evidence E, and E confirms P, then you are not in a position to rule out P after receiving

evidence E there's hallucination gas that undermines any knowledge that would allow me to rule out E and not-P, but the hallucination gas has dissipated by the time I make the inference, then my inference that P may in fact constitute knowledge.

[23] Principles with this general form are called reflection principles; see Fraassen (1984).

[24] Presumably if inductive knowledge is possible, we can know it is. So presumably there is some suitably large number — a billion say — such that we know that that many observations of instances of the law suffices for knowledge that the law holds.

evidence E.

Here E confirms P just in case P is more likely conditional on E than unconditionally, and by 'ruling out P' I just mean 'knowing not-P'. The bonus knowledge theory predicts failures of this principle as well. Initially we should have very little credence that all the emeralds e_1 to e_N are green by fluke. (Assuming that conditional on being random the probability of being green is $\frac{1}{2}$, and independent for each emerald, then the probability that they are all green by fluke will be less than $\frac{1}{2}^N$!) But conditional on the fact that $e_1...e_N$ *are* all green, it should be much more likely (by orders of magnitude)[25] that they are all green by fluke than it was initially. Indeed, one would have thought that after learning that $e_1...e_N$ are green we should assign a greater amount of confidence to the possibility that they were all green by fluke. If we were not in a position to rule out that possibility before we received that evidence we certainly shouldn't be in a position to rule it out afterwards.

It is also worth noting that, assuming the bonus knowledge theory, there are strong reasons to think that we can know that the first N emeralds won't all be green by fluke before investigation. If this is right then it is not obvious why we need the bonus knowledge theory — we already have the prior knowledge needed to explain how we achieve knowledge by induction. Let us begin with the view that the tipping point occurs after a relatively stable number of observations: for the sake of argument suppose it is 100. Then it seems we could be in a position to know that 100 was the tipping point (perhaps, e.g., by studying how long it takes others to knowledgeably conclude that it's a law that emeralds are green).[26] To know the tipping point is 100 is just to know, before investigation, that the following conditional is true:

> If I observe $e_1...e_{100}$ to be green I will be in a position to know that it's a law that emeralds are green.

Suppose in addition you know that you're going to observe $e_1...e_{100}$, so that you know that you'll observe e_i to be green if it is in fact green (and know that you'll observe e_i to be blue if it is in fact blue). Given closure, I can infer that if $e_1...e_{100}$ are green I'll be in a position to know it's a law that emeralds are green. Finally by applying the factivity of 'in a position to know', I'll be able to conclude that if $e_1...e_{100}$ are green it's a law that emeralds are green; in other words, I'll know that $e_1...e_{100}$ won't be green by fluke.[27]

Of course, it could be that the tipping point is very variable and that people are rarely in a position to know when it will occur. But if we are in a world where inductive knowledge is in practice possible after a reasonable number of

[25] The difference in probability itself may not be that large: the difference between, e.g., $\frac{1}{2}^{100}$ and $\frac{1}{2}^{200}$ is small even though the former is billions of times bigger than the latter.

[26] Of course, one could be unlucky: one could inhabit a world where the first 100 emeralds everyone observes is green, but there are nonetheless blue emeralds. In these worlds we would not be in a position to know the tipping point (there wouldn't be one), but these quasi-skeptical scenarios should not undermine the possibility of our actually knowing where the tipping point is.

[27] Again, the problem here mirrors a similar worry for the dogmatist; see Wedgwood (2013).

observations then it seems we ought to be in a position to know that inductive knowledge is in practice possible. Thus there ought to be a not too large number, N, possibly greater than the tipping point, such that we know that we'll be in position to know that emeralds are green after observing that many emeralds. Thus, as above, we will be in a position, prior to investigation to know that a certain sequence of emeralds will not be all green by fluke.

An initially attractive feature of the bonus knowledge theory — one of the strongest reasons to prefer it to the prior knowledge theory — is that it appears to avoid the postulation of this mysterious sort of knowledge that can be obtained prior to an empirical investigation. If the forgoing considerations are correct, then the bonus knowledge theory is actually not in as good a position as it seems.

The bonus knowledge theorist might at this point insist that we cannot know when the tipping point will occur until we've actually reached the tipping point.[28] They could maintain that, since for all we know we are in the bad case where it is not a law but $e_1 ... e_{100}$ are green by fluke, we couldn't possibly know that if we made those observations we would be be in a position to know that it was a law. The most we can know is that in the good cases, where it is in fact a law that emeralds are green, we can know that emeralds are green by law on the basis of observing $e_1 ... e_{100}$ to be green. That is, we can know the weaker conditional:

> If it is *in fact* a law that emeralds are green, and we observe that emeralds $e_1 ... e_{100}$ are green, we will be in a position to know that it's a law that emeralds are green.

The picture seems to be one in which we cannot know that inductive knowledge is actually possible until we've achieved it. Of course, the weaker conditional tells us that we can know that *in the good cases* where it is a law that emeralds are green we can achieve knowledge by induction, but this is silent on the actual status of inductive knowledge. But why is it that we can't know the original stronger conditional before looking at any emeralds? It is hard to see why we need to observe any emeralds to acquire knowledge of the conditional; it seems to be something one could plausibly know in a myriad of different ways. Perhaps we just know it innately, or we know it by 'metainduction': i.e. by observing that previous inductive inferences produce knowledge (not merely that they produce knowledge in the good cases).

4 Puzzles for the Prior Knowledge Theory

We just observed that, under certain assumptions, the bonus knowledge theory collapses into some form of the prior knowledge theory. In this section we shall raise some puzzles for the prior knowledge theory. We show first that it conflicts with a natural independence idea, which roughly says: if it's consistent

[28] Thanks to an anonymous referee for pushing me on this point.

with your knowledge that some emeralds are green by fluke, and that some other emeralds are green by fluke, then it should be consistent that they are all green by fluke. Then we show, assuming multi-premise closure, that it generates problems when combined with a natural idea which I'll refer to as 'emerald anonymity'[29]: roughly, the idea that observing that a particular emerald is green should be as informative as observing that any other emerald is green.

4.1 Epistemic Independence

Suppose that after observing that all the emeralds in a given set, X, are green, Alice is still not in a position to know whether it's a law that emeralds are green. She then boldly makes the following assertion:

(A) If it's *not* a law that emeralds are green, then the next emerald will be blue.[30]

This seems like a risky assertion. It would be very perplexing if Alice were indeed in a position to know (A): how could she possibly know that if the emerald's colors are determined by chance, the next emerald is going to be blue?

In order to rule out knowledge of (A) in this circumstance we must thus accept:

Epistemic Independence 1 If, prior to investigation, it was epistemically possible that the emeralds in X are all green by fluke, then it is epistemically possible that the emeralds in X and e are green by fluke, where e is any emerald not already in X.

For suppose that Epistemic Independence 1 has a false instance. That is to say: suppose that prior to investigation it were epistemically possible that the X emeralds are all green by fluke, but not epistemically possible that the X emeralds and the ith emerald are green by fluke. Then the claim that the X-emeralds are green and the prior knowledge entail that either we are in a flukey world and the ith emerald is blue, or we are in a lawlike world and all emeralds are green. So if Alice learns that the emeralds in X are all green, she should be in a position to know, and thus assert (A), contrary to our intuitions otherwise.

I take it that the following parallel inference is equally perplexing. After observing the elements of a set Y to all be green, Bob is still not in a position to know that it's a law that emeralds are green. But nonetheless he asserts:

(B) If it's not a law that all emeralds are green then the next emerald will be green by fluke.

Again, this assertion seems bad because Bob does not seem to be in a position to know (B).[31] To rule out this sort of knowledge, we need to accept the following

[29] Thanks to Shyam Nair for suggesting this terminology

[30] This is to be understood as material conditional, so it is equivalently: either it's a law that all emeralds are green or it's not a law and the next emerald will be blue.

[31] Note that (A) and (B) are not inconsistent: they are both material conditionals with false antecedents. Note also that, since X and Y are different sets of emeralds, we have no reason to think anyone is will be in a position to know both conditionals at once.

principle:

Epistemic Independence 2 If, prior to investigation, it's epistemically possible that the emeralds in X are all green by fluke, then it is epistemically possible that these emeralds are green, and that e is blue by fluke, where e is any emerald not already in X.

Epistemic Independence 1 sounds eminently plausible. Moreover, (A) — whose assertability is a consequence of its failure — sounds particularly bad. If I don't know yet whether it's a law that emeralds are green, it seems quite absurd to think that I'd be in a position to know what colour the next emerald will be if it's not a law that emeralds are green.

This makes the following somewhat puzzling for the prior knowledge theory:

Skeptical Puzzle 2 The following are inconsistent

1. Prior to investigation it's true for all I know that e_1 is green even though there is not a general law that emeralds are green.
2. It is known, prior to investigation, that $e_1 ... e_N$ are not all green by fluke.
3. Epistemic Independence 1

We shall show that 1 and 3 entail that it's epistemically open prior to investigation that the emeralds are all green by fluke, meaning that the explanation of inductive knowledge proposed by the prior knowledge theory cannot be right. Since, prior to investigation, it is open to me that the first emerald is green but it's not a law that emeralds are green, it follows by Epistemic Independence 1 that it's epistemically possible the first two emeralds are green by fluke. By Epistemic Independence 1 again I can infer that it's open to me that the first three emeralds are green by fluke, and by repeating this N times I can infer that it's epistemically open prior to investigation that all N observed emeralds are green by fluke, directly contradicting prior knowledge.

As far as I can tell, there is no straightforward argument against Epistemic Independence 2 from prior knowledge. However the specific versions of the prior knowledge theory considered later involve failures of Epistemic Independence 2 as well. Since Epistemic Independence 1 must already be rejected given the prior knowledge theory of induction, it is hard to see what motivation one could have for insisting on Epistemic Independence 2.

It's worth mentioning that Dorr et al. (2014) argue for similar puzzles without invoking any premises about the correct theory of induction. Given that Epistemic Independence 1 and 2 follow from more general independence intuitions that are false, we should be cautious about concluding much from the fact that a theory entails failures of Independence.

4.2 Emerald Anonymity

To motivate the next principle it's easiest to illustrate it with an example. Suppose that 100 emeralds are laid out on a table and that after observing each

one to be green you are in a position to know that it's a law that emeralds are green. Here's a very natural intuition: if prior to observation, someone had switched each emerald on the table for a different emerald, and that the observations had otherwise proceeded exactly as before, you would still be in a position to know that it's a law that emeralds are green after making those observations.

The principal guiding the intuition here is that it shouldn't matter *which particular emeralds* you observe; the observations should have the same evidential impact. Other things being equal, the fact that one emerald is green is just as informative to you as the fact that any other emerald is. Let's give this principle a name.

Emerald Anonymity If you're in a position to know that it's a law that all emeralds are green after observing that emeralds $e_1 ... e_n$ are all green, one should also be in a position to know this after observing $\pi e_1 ... \pi e_n$ to be green, where π is any permutation of the 1000 emeralds.

This principle has some initial plausibility, but one must be diligent about when it is and isn't applicable. For example, suppose, prior to investigation, you have been told by a reliable person that either the 23rd emerald is not green or it's a law that emeralds are green. Then it's clear that not all sets of emeralds of the same size are on a par in the way that Emerald Anonymity demands. If you observe a collection of three emeralds to be green you will be in a position to know that it's a law that emeralds are green if it includes the 23rd emerald, but not otherwise. The principle is not supposed to apply to cases like this, where someone has specific knowledge about the emeralds.

Given Emerald Anonymity it follows that, for each agent, there is a *magic number* of observations sufficient for inductive knowledge: the smallest number of emeralds they need to observe to be green in order to conclude that it's a law that emeralds are green. This number will not depend on which emeralds are observed. For if I can know that it's a law that emeralds are green after observing n distinct emeralds $e_1 ... e_n$, then I must also be in a position to know this after observing any other series of n distinct emeralds $e'_1 ... e'_n$ by applying Emerald Anonymity to a permutation that maps e_i to e'_i.

Let me finally mention another plausible idea. Clearly we cannot know, before looking at any emeralds, that it's a law that emeralds are green. Nor can we know, before looking at any emeralds, that if it's not a law that emeralds are green, then 90% of the emeralds will be blue. Thus we may state our third skeptical puzzle. Suppose that 100 is the magic number for some particular agent. Recall also that in the present set up there are a total of 1000 emeralds, and that this is known.

Skeptical Puzzle 3 Inductive Knowledge, Multipremise Closure, What You See is What You Know, and Emerald Anonymity entail Too Much Prior Knowledge.

For convenience, here are the relevant principles restated:

Inductive Knowledge After observing $e_1...e_{100}$ to be green we are in a position to know that it is a law that all emeralds are green.

Multipremise Closure: If you know $p_1...p_n$ individually, and they jointly entail q, then your are in a position to knowledgeably infer q.[32]

What You See is What You Know: Your knowledge after looking at emerald e_k is just whatever your knowledge was before plus the proposition that e_k is green.

Emerald Anonymity: If π is a permutation of all 1000 emeralds, then if you're in a position to know that it's a law that emeralds are green after learning $e_1...e_n$, you are in a position to know this after observing $\pi e_1...\pi e_n$.

Too Much Prior Knowledge: Prior to observing any emeralds you are in a position to know the material conditional that if it's *not* a law that all emeralds are green, then at least 90% of the emeralds are blue.[33]

We shall break this argument up into three steps. Firstly we use Inductive Knowledge, Closure and What You See is What You Know to show:

> Prior to observation you know the material conditional that if $e_1...e_{100}$ are green then it's a law that emeralds are green.

In fact we have already sketched an argument for this in the introduction, but here we set out the argument more carefully.

We may argue by contradiction. Suppose that you do not know this conditional prior to observation: it's true for all you know that $e_1...e_{100}$ are green, and it's not a law that emeralds are green. By closure this means that this conditional does not logically follow from things you know prior to observation. By What You See is What You Know and Closure, you're in a position to know a proposition after making the observations if it can be logically deduced from propositions you knew initially along with the new knowledge consisting of the propositions that $e_1...e_{100}$ are green. But if a conditional $(P_1 \wedge ... \wedge P_k) \rightarrow Q$ cannot be logically deduced from a set of propositions S, then Q cannot be logically deduced from $S \cup \{P_1, ..., P_k\}$.[34] In particular, if you can't deduce the conditional *if $e_1...e_{100}$ are green then it's a law that emeralds are green* from propositions you know initially, then you cannot deduce the consequent *it's a law that emeralds are green* from propositions you know after making the observations. Thus you do not know that it's a law that emeralds are green after making the observations. This contradicts Inductive Knowledge.

Secondly, using Emerald Anonymity, the above argument can be repeated to establish that prior to observation you know the material conditional that if

[32]Note that the special case of this with 0 premises implies that the agent knows every logical truth.

[33]Equivalently: you are in a position to know, prior to investigation, that either it's a law that all emeralds are green or 90% of the emeralds will be blue.

[34]This is just a version of the deduction theorem.

$\pi e_1 ... \pi e_{100}$ are green then it's a law that emeralds are green, for any permutation π of $e_1 ... e_{1000}$. This is clearly equivalent to:

> Prior to observation you know the material conditional that if $e_{i_1} ... e_{i_{100}}$ are green then it's a law that emeralds are green.

for any distinct emeralds $e_{i_1} ... e_{i_{100}}$ (simply consider a permutation that maps k to i_k for $1 \leq k \leq 100$).

Finally, by Multipremise Closure, we know the conjunction of these conditionals. A conjunction of conditionals $(P_1 \rightarrow Q) \wedge ... \wedge (P_k \rightarrow Q)$ is equivalent to $((P_1 \vee ... \vee P_k) \rightarrow Q)$. In the present case: if either $e_{i_1} ... e_{i_{100}}$ are green, or $e_{i'_1} ... e_{i'_{100}}$ are green or ... then it's a law that all emeralds are green (where $(i_k), (i'_k)$ etc range over all sequences of length 100 consisting of numbers between 1 and 1000 without repetitions). Recall that our background knowledge included the fact that $e_1 ... e_{1000}$ where all and only the emeralds in existence. Thus this long disjunction in the antecedent is fairly obviously equivalent, given the background knowledge, to the claim: if 100 or more emeralds are green then it's a law that emeralds are green. Or equivalently (again, given the background knowledge): if 10% of the emeralds are green then it's a law that emeralds are green. By closure, then, we know this conditional. And by closure again, we also know its contrapositive:

> Prior to observation you know the material conditional that if it's not a law that emeralds are green, 90% of the emeralds are blue.

(We may again illustrate the result in a simple possible worlds model of knowledge, in which one's knowledge at a time is represented by a set of worlds and a proposition is known if it is a superset of that set. Suppose, for contradiction, that a set of worlds K represents my prior knowledge, and that there is a world $w \in K$ in which less than 90% of the emeralds are blue but also it's not a law that emeralds are green or blue. Thus in w, some sequence $e_{i_1} ... e_{i_{100}}$ of 100 (possibly more) distinct emeralds are green and it is moreover not a law that emeralds are green in this world: $w \in G_{e_{i_1}}, ..., G_{e_{i_{100}}}$ and $w \notin LAW$ where LAW is the set of worlds in which it's a law that emeralds are green and G_i is the set of worlds where e_i is green. It follows by Inductive Knowledge, that $K' \subseteq LAW$ where K' is my knowledge after observing $e_1 ... e_{100}$. Then, by Emerald Anonymity, we are in a position to know that it's a law that all emeralds are green after observing the emeralds $e_{i_1} ... e_{i_{100}}$ (consider the permutation that maps 1 to i_1, etc): if K'' is my knowledge after observing $e_{i_1} ... e_{i_{100}}$ then $K'' \subseteq LAW$. But given What You See is What You Know, $K'' = K \cap G_{e_{i_1}} \cap ... \cap G_{e_{i_{100}}}$. So $K \cap G_{e_{i_1}} \cap ... \cap G_{e_{i_{100}}} \subseteq LAW$. This is a contradiction since $w \in K$ and $w \in G_{e_{i_1}}, ..., G_{e_{i_{100}}}$, but $w \notin LAW$.)

The numbers 100 and 1000 were chosen for the sake of illustration: whatever the magic number is, one could imagine a world where there are known to be ten times that many emeralds and arrive at the same conclusion.

One might try to reject the previous argument by maintaining that an agent's magic number depends on how many emeralds there are. One might maintain,

for example, that the magic number is always 60% of the total number of emeralds. Since it is statistically abnormal for there to be less than 40% blue emeralds (assuming blue and green is equally likely for each emerald) perhaps we have a good claim to knowing, at least in the good, statistically normal cases, that if it's not a law, then at least 40% will be blue. In this case, the conclusion seems to be less obviously absurd. However even though this view is consistent with the letter of Inductive Knowledge, it doesn't seem particularly friendly to actual inductive knowledge, since it turns out one has to have observed most of the emeralds (over 60% of them) before one can get obtain knowledge. This is just not good enough for science: compare, for example, the number of observed electrons to the number of electrons in the universe simpliciter — if laws about electrons can be known by induction it has to be based on observations of a tiny fraction of the total number of electrons.

The conclusion we have arrived at seems to be obviously false: that one can know prior to investigation that most emeralds will not be green, if it's not a law that they're green. One might hope to lessen the force of the objection by noting that I can't learn that it's not a law that all emeralds are green (because it is in fact a law, and knowledge is factive), and thus I will never be in a position to infer, before observation, that 90% of the emeralds will be blue from the conditional. But this seems to me to be just to point out that we can avoid an even more absurd conclusion; knowledge of the conditional should strike us as bad enough.

In light of this, I think the best option for the prior knowledge theorist is to reject Emerald Anonymity. The challenge, then, is to make this rejection seem motivated: there is an appearance of symmetry between the different emeralds — if we are to maintain that this symmetry is broken, we must identify an epistemically significant difference between the emeralds that could ground that failure of symmetry.

Before we move on, it's worth remarking that there is no parallel argument that the bonus knowledge theory must reject the analogue of Emerald Anonymity: for all we've said, it could just be that the tipping point is reached at 100 emeralds no matter which particular emeralds are observed.

5 Counterfactuals

There is a large body of theory on the purported relation between counterfactuals and knowledge. Here I will draw on a connection that has been mostly been discussed in the literature on safety.

At a rough gloss, a belief is *safe* if there is no nearby world at which it is had falsely, for the same reasons it is actually had. We will follow the literature in using the locution *it could easily have been the case that* P when there is a nearby P world. According to a very compelling thought, safety is a necessary condition for knowledge. If you could easily have had the belief that P for the same reason you actually had it, while P is false, then it seems that even if P is in fact true, your belief that P couldn't count as knowledge: even though you

were right, you were just too lucky to count as having known.

This appeal to *nearby worlds* is essentially just a helpful heuristic: the notion of being nearby is ultimately to be understood, at least partly, in epistemic terms, and cannot obviously be given an independent analysis.[35] What could easily have been the case for me at t may differ from what could easily be the case for you at t, and may differ from what could easily be the case for me at other times; these differences depend on what can only be described as *epistemic* features of the situations.[36] Nonetheless, there may be connections between this concept and other non-epistemic concepts that help pin down its conceptual role. Here is one inference that, when suitably qualified, I find particularly compelling:

1. It could easily have been the case that P.
2. If P were the case then Q would be the case.
3. Therefore, it could easily have been the case that P and Q.

I take it that counterfactuals are context sensitive, and as noted above, that what could easily be the case for an agent at a time can depend on what that agent is already able to rule out at that time. Given this fact we should not expect this inference to hold in all contexts for what could easily have been the case for all agents and times.[37] That said, it's natural to think that there are certain contexts in which the inference is a good one, and in these contexts the English counterfactual will express a special sort of connective that is sensitive to safety theoretic concerns of the agent and time in question.

One way to motivate the inference 1-3 is through the Lewis-Stalnaker style of semantics for counterfactuals, in which a counterfactual is true iff the nearest antecedent world to actuality is a consequent world.[38] Here the notion of nearness is context sensitive, but in certain contexts the notion of nearness will align with the safety theoretic notion of nearness invoked in 1 and 3. Suppose there's a nearby world, w, in which P is true, and also that the nearest P world, w', is a Q world. It follows that w' is at least as near as w, and so w' certainly also counts as nearby if w does. Thus there is a nearby P and Q world.

[35] See Williamson (2009). There is a spectrum of positions here: Williamson's view seems to be that the utility of the safety model of knowledge lies primarily in the structural constraints it imposes on the concept of knowledge, and that there is no definition of nearness in non-epistemic terms. Others attempt to explain it in terms of counterfactual nearness (see Sosa (1999)), and there are intermediate positions too. I needn't take a position on this in what follows.

[36] 'It could easily be the case P' therefore could mean different things depending on the agent or the time; we will use context to disambiguate in what follows.

[37] Given that knowledge ascriptions are plausibly context sensitive, it is perhaps also true that the notion of 'being nearby' is context sensitive. One could thus maintain that meanings of counterfactual and nearness locutions vary from context to context in tandem, preserving the truth of the inference. But I am suspicious of this further claim, so will leave it aside in what follows.

[38] I am working with Stalnaker's version of the theory. Lewis has some tweaks to deal with cases in which there is more than one nearest antecedent world, or no nearest antecedent world (Stalnaker rules out such cases by fiat).

That said, even if one does not accept the Lewis-Stalnaker semantics, the principle seems independently quite compelling. To illustrate, suppose that I correctly believe that Fiona is happy. Unbeknownst to me, she has had a rotten week but she's just learned that she has won the lottery and so, as it happens, is happy. It seems as though my true belief that Fiona is happy was too lucky to count as knowledge: it could easily have been false. I take it we can substantiate this conclusion as follows: the lottery numbers could have easily been different, and if they had been different Fiona would not have been happy. Thus my belief that Fiona is happy could easily have been false.

We must also be cautious about applying the inference, as it is quite easy to get into contexts where the inference does not sound so good. Suppose that you have two buttons before you, and I know the following: (i) that one of the buttons, I don't know which, will kill me, (ii) that you have to press a button, know which button does what, and do not want to kill me. Given that I know you won't kill me, I know:

> If you were to press button 1, I wouldn't die.

According to the epistemic reading of the counterfactual, the above counterfactual should be true, as it indeed seems to be in this context. But of course, the buttons may in fact have been wired so that pressing button 1 would kill me. In a context where this fact is emphasized, it's tempting to want to assert instead that:

> If you were to press button 1, I would die.

In this context the inference is bad: For all I know, you'll press button 1. As it happens, if you were to push button 1 I would die. However, contrary to our inference, it's not the case that for all I know you'll kill me by pressing button 1. There is often a strong prejudice to keep the functioning of mechanisms, like buttons, fixed when evaluating which worlds are relevant to the truth of a counterfactual.[39] But in this case, this default constraint conflicts with the epistemic understanding of the counterfactual: the epistemically nearest world where button 1 is pushed is one where button 1 is wired differently, and is the button that spares me.[40]

6 Counterfactuals and Inductive Knowledge

As a warm up to our discussion of induction, let us apply these ideas to a couple of more quotidian cases. Suppose that I know that my keys are either in the house or in the car. As it happens they are in the car. Now I take it that one does not usually need to have checked every cubic centimeter of the house in

[39] More generally, there's a prejudice to keep the laws of physics held fixed, although it's widely believed that this cannot hold in full generality. The idea is especially problematic if those laws are deterministic, so that any deviation from reality at a time would result in a deviation at all times, future and past, unless the laws were broken.

[40] Thanks to John Hawthorne for this example.

order to be able to know that the keys are in the car. In good cases, we can acquire knowledge by this method:

Lost Keys (Good Case) My keys would have been on the mantlepiece or the table (as usual) if they had been in the house. After checking a few of the obvious places, including the mantlepiece and table, I get to know the keys are in the car.

I think the analogy with inductive knowledge here should be evident: I do not have to observe every emerald in order to know that all emeralds are green, just as I do not need to search the whole house in order to know that my keys are in the car. But there are also cases where this way of acquiring knowledge doesn't work:

Lost Keys (Bad Case) My daughter has developed a fascination with my keys, and likes to throw them down the back of the sofa. My keys would have been behind the sofa had they been in the house.

In this case, I take it, I do not get to know that my keys are in the car after checking the table, mantlepiece and other obvious places. The reason is that my keys wouldn't have been in those places if they hadn't been in the car: they would have been behind the sofa. One can substantiate this conclusion with the inference 1-3 as follows. It is certainly true, prior to searching, that the keys could easily have been in the house.[41] Moreover, by stipulation, if the keys had been in the house, they would have been behind the sofa. Thus by the counterfactual inference, the keys could easily have been behind the sofa. On the other hand, if I had checked behind the sofa, and perhaps other similar places my daughter might have placed them had the keys been in the house, then it seems that I could be in a position to know.

To connect this to the literature on dogmatism, we might also consider a perceptual case. Suppose that at a flip of a coin you will either be shown a red ball or a white ball. You are shown a ball that looks red, and is in fact red. We assume that the experiment is set up so that you would have been shown the white ball in ordinary lighting, had the white ball been shown instead.

Perception (Good Case) You get to know that the ball is red. For had the ball been white, it wouldn't have looked red, it would have looked white.

But there is also a version of the experiment where you see a ball that looks red and is red, but you intuitively don't get to know that it's red:

Perception (Bad Case) The set up is changed so that had you been shown the white ball, it would have been under red lighting. Even though the ball seems red, and is red, intuitively we don't know.

[41] It's worth noting that this intuition, and the ones that follow, could also be underwritten by a sensitivity account of knowledge. Suppose that in order for me to know a proposition P by a method M the following counterfactual must be true: had P been false, I wouldn't have believed P by the method M. Then I do not know that the keys are in the car by checking the table and mantlepiece: for had the keys not been in the car, I would still have believed they were in the car by the method of checking the table and mantlepiece. By contrast, the method of checking the behind the sofa does much better in this regard.

The reason we don't know, I submit, is that if the ball had been white it would have looked the same as the red ball.[42] Again, this verdict is underwritten by the inference 1-3: since prior to observation, the ball I was shown could easily have been white, and had it been white, it would have looked red, it follows that it could easily have been white and looked red.

These sorts of examples reveal a source of counterexamples to Emerald Anonymity, a principle which played an important role in the puzzle we discussed in section 4.2.[43] Suppose that millions of years ago, a team of alien scientists landed in New Mexico and performed a number of geological experiments. In the course of these experiments, they ran a test which vaporized all mineral crystals in New Mexico except those that are green in color. Now imagine two modern day scientists, Anna and Beth, and suppose that Anna is based in California while Beth is based in New Mexico. It is in fact a law that emeralds are green, and both Anna and Beth go about observing emeralds in their respective locations and discover that every emerald they observe is green: Anna observes 100 Californian emeralds to be green, call these $e_1 ... e_{100}$, and Beth observes 100 green New Mexican emeralds, $e'_1 ... e'_{100}$. Even though neither Anna nor Beth know about the alien landing, there appears to be an epistemic asymmetry between the cases: it seems like Anna should be in a position to know that it's a law that emeralds are green (assuming 100 is a sufficient number of observations), whereas Beth intuitively is not, since she would have observed those emeralds to be green *even if it hadn't been a law.*

Californian Emeralds (Good Case) On the basis of observing emeralds $e_1 ... e_{100}$ (the Californian emeralds) to be green, one is in a position to know that emeralds are green by law.

New Mexican Emeralds (Bad Case) On the basis of observing emeralds $e'_1 ... e'_{100}$ (the New Mexican emeralds) to be green, one is not in a position to know that emeralds are green by law, since even if it hadn't been a law, $e'_1 ... e'_{100}$ would have been green anyway.

Clearly, there's a symmetry between Anna's evidence about $e_1 ... e_{100}$ and Beth's evidence about $e'_1 ... e'_{100}$ (neither know about the alien landing). So this appears to be a counterexample to Emerald Anonymity.

If this is right, that means straightforwardly probabilistic accounts of the tipping point are not on the table. For example, the theory that you are in

[42] It should be noted that certain externalists might seek to maintain that when we look at the red ball, even if there would have been trick lighting had the ball been white, we still get to know that the ball is red. The anti-dogmatist version of this view maintains that if you're going to be shown the red ball, then prior to the showing you know the material conditional: if the ball looks red, it is red. But this material conditional entails, given the setup, the outcome of an unknown coin toss (since the choice between being shown a red ball or white ball in trick lighting was determined by a coin toss). Thus (given closure) one could knowledgeably conclude how the coin landed, which strikes me as absurd. The dogmatist version of the view, where you get to infer the outcome of the coin toss upon seeing a ball that looks red seems to me to be only marginally better.

[43] Thanks to an anonymous referee for suggesting a similar sort of example.

a position to know something when it becomes sufficiently probable on your evidence (i.e. its probability exceeds some threshold) is a version of the bonus knowledge theory, since the observation that e_N is green can bring other previously unknown propositions over the threshold (including the proposition that it's a law that emeralds are green). But because the tipping point does not depend on what counterfactuals are true on this theory it will not predict the variance across worlds we have demonstrated above.[44]

Another moral to draw from our considerations about counterfactuals is that the point at which we get to know that emeralds are green by law can depend on *which emeralds we observe*, revealing a source of counterexamples to Emerald Anonymity. This phenomenon is brought out in our above example. If that intuition is correct, Californian emeralds are more informative that New Mexican emeralds: if the emeralds you are observing contain a large number of New Mexican emeralds you need to make more observations than you would otherwise (and if they are solely New Mexican emeralds, no amount of observations will do).

However, while our example illustrates a somewhat exotic exception to Emerald Anonymity, it is not obvious that it does anything to deflate the paradox of section 4.2. For while it is clear what grounds the asymmetry between Californian and New Mexican emeralds in the strange world described earlier, the upshot of the puzzle of section 4.2 was that Emerald Anonymity must have actual failures. However, it is completely unclear what could ground the difference between one collection of emeralds and another of the same size, given that we have not in fact been visited by aliens nor, we may assume, has anything similar happened to force the counterfactual facts to be one way or another.

The way this sort of worry is often formulated makes the illicit assumption that when some symmetry is broken, there has to be some 'tangible' fact that grounds the breaking of that symmetry. But there are theories of counterfactuals in which counterfactuals like (*) below could just be true by chance (even though it is incredibly improbable):

(*) If it hadn't been a law that emeralds are green, emeralds $e_1 ... e_{100}$ would have been green anyway.

Indeed, this could happen even while the counterpart counterfactual (**) is false (again by chance):

(**) If it hadn't been a law that emeralds are green, emeralds $e_{101} ... e_{200}$ would have been green anyway.

This, of course, in itself breaks a symmetry between the emeralds $e_1 ... e_{100}$ and the emeralds $e_{101} ... e_{200}$. It is, moreover, a fallacy, according to these sorts of

[44]This argument assumes that what your evidence consists in is not sensitive to counterfactuals in the same way that knowledge is. If one assumes, along with Williamson (2000), that your evidence is your knowledge, and you also think that your knowledge is sensitive to what counterfactuals are true then one could resist this argument. I have argued elsewhere against Williamson's theory of evidence; see Bacon (2014).

theories, to think that there must some non-counterfactual, categorical fact that distinguishes the first collection of a hundred emeralds from the second.

Note, however, that regardless of whether (*) is true by chance or because of alien interference: if it is indeed true, then we can infer (using (1)-(3)) that we are not in a position to know that emerald $e_1...e_{100}$ aren't green by fluke prior to observation. For if, prior to investigation, we're not in a position to know that it's a law that emeralds are green — a belief in this proposition could easily have been false — then it follows from (*) and the counterfactual inference (1)-(3) that it could easily have been the case that the first 100 emeralds are all green by fluke. (More precisely: it follows that it could easily have been the case that $e_1...e_{100}$ all green, and that it is not a law that they are all green).

Note, on the other hand, that if (**) is false, there is no analogous sound argument that we are not in a position to know that $e_{101}...e_{200}$ will all be green by fluke. Now one can see how the asymmetries predicted by failures of Emerald Anonymity can be predicted by the asymmetries concerning what colours the emeralds *would have had* had it not been a law that emeralds are green.

This answers our puzzle. What is breaking the apparent epistemic symmetry between the observations of $e_1...e_{100}$ and the observations of $e_{101}...e_{200}$ is a difference in the *counterfactual* facts: by facts about what the emeralds colours would have been, had the laws been different. These differences are, admittedly, invisible: you can't tell by looking at an emerald what colour it would have been had the laws been different, you can only tell what its actual colour is. But presumably it is the invisibility of the differences between the emeralds that makes the intuition that there is no relevant epistemic difference, and consequently the intuition for Emerald Anonymity, so inviting.[45]

It's worth noting that the same could be said about the lost keys case. There's an 'invisible' difference between places: the mantlepiece and table are better places to look in the first scenario, whereas the back of the sofa is a better place to look in the latter. What grounds these differences isn't some visible difference between the sofa and the table and mantlepiece; it is the counterfactual facts that single out the sofa as a better (more knowledge conducive) place to look in the latter situation, than elsewhere. So we shouldn't be too surprised that epistemic differences like this can be grounded by counterfactuals. What's surprising about the induction case is it's unclear what grounds the differences in the *counterfactual* facts: in the key case, at least, the relevant counterfactuals are grounded by my daughter's fascination with hiding shiny objects. There is nothing analogous grounding the counterfactual 'if it hadn't been a law, emerald e would have been green' (at least, assuming that we have not been visited by aliens, or what have you). It is a bit like asking what grounds the true disjunct of the disjunction 'the coin either would have landed heads if it had

[45] It's also important to emphasize that questions like 'what colour would this have been had the laws been different?' are incredibly context sensitive and vague. Note however, that we are constraining the context sensitivity by requiring the counterfactual be a 'safety-theoretic' counterfactual. And even if the question is vague, vague questions still have answers. For, assuming classical logic, either this emerald would have had a particular colour, had the laws been different, or it wouldn't.

been flipped, or it would have landed tails' said of an unflipped coin. When the counterfactuals in question are chancy or indeterminate — as they plausibly are in these cases — it is a mistake to think that they must be grounded by some identifiable non-counterfactual fact.

The morals we drew above rested on the assumption that there were worlds in which counterfactuals like (*) are true by chance. According to one picture, worlds in which the relevant counterfactuals are true are relatively easy to come by. This is a view which maintains the principle of *conditional excluded middle* is valid for counterfactuals. According to that view the counterfactual *if P were the case then Q or R would be the case* entails the disjunction *either Q would be the case if P or R would be the case if P*. In symbols: $P \Box\!\!\rightarrow (Q \vee R) \vdash (P \Box\!\!\rightarrow Q) \vee (P \Box\!\!\rightarrow R)$. It follows, given our setup, that for any $1 \leq i \leq 100$:[46]

> Either e_i would be green, if it weren't a law that emeralds were green, or it would be blue if it weren't a law that emeralds are green.

Given the symmetry between the two disjuncts, it's natural to think that roughly one in two worlds are worlds where e_i would have been green, and the other one of those two worlds is a world where it would have been blue. Worlds where $e_1 ... e_{100}$ would all have been green, had there not been a law, are presumably much more rare, but exist by similar sorts of considerations.[47] The validity of conditional excluded middle is a vexed topic that I do not wish to get into. But even without it, the point that the possibility of inductive knowledge depends on a hospitable environment of counterfactual facts still remains. There are many theories concerning what it takes to make a counterfactual true; a particularly demanding account requires that the laws along with the antecedent entail the consequent. But the most plausible accounts of counterfactuals are not this demanding, and it's not unreasonable to think that on these more liberal sorts of theories one could concoct cases where things are aligned in the correct way so as to make the counterfactual *if it hadn't been a law that emeralds are green, $e_1 ... e_{100}$ would have been green anyway* true.

One might wonder how commonplace worlds which are inhospitable to inductive knowledge are. If they are ubiquitous then we should be worried about whether the actual world is one in which inductive knowledge is possible. Extending the informal reasoning we gave above, it's natural to think that one in 2^{100} worlds are worlds where $e_1 ... e_{100}$ would have been green if it hadn't been a law that emeralds are green. For given conditional excluded middle type reasoning, at every world there is some assignment of colours to $e_1 ... e_{100}$ that

[46] This is a consequence given our assumption that if it weren't a law that emeralds are green, then emeralds can either be green or blue. A longer (potentially infinite) disjunction would be necessary if this assumption weren't made, but the point would be essentially the same.

[47] It follows from the set up that *if it hadn't been a law that emeralds are green, then either $e_1 ... e_{100}$ would have distribution d_1, or d_2 or d_3 or ...* where $d_1, d_2 ...$ range over the 2^{100} possible distributions of colours $e_1 ... e_{100}$ could have. One can then infer from this a disjunction of counterfactuals that has 2^{100} disjuncts, each disjunct of which is presumably true in some worlds.

represents the colours they would have had, had it not been a law that emeralds are green. There are, moreover, 2^{100} assignments. Absent any reason to think certain counterfactual assignments are more common we should assume that any particular counterfactual assignment occurs infrequently.[48] Thus even though there *are* worlds where induction is not possible, they are sufficiently rare that we might treat them as quasi-skeptical scenarios: although there are unlucky people who fail to get knowledge from the same observations we make, these people are in the same boat as brains-in-vats who fail to get knowledge by perception — these scenarios have no bearing on the actual efficacy of induction or perception. (A perhaps closer analogy: although there are worlds where my bank account has been hacked by someone who has correctly guessed my ten digit PIN, it is tempting to think that I am in fact in a position to know that my account is secure from attacks by lucky PIN-guessers, and that the envisioned scenario is also a quasi-skeptical scenario.)

It's worth stressing that when we talk about some worlds being 'unlucky' in the sense that the counterfactuals are inhospitable to inductive knowledge, we mean that we are unlucky holding fixed the fact that $e_1...e_{100}$ are being observed first. In most worlds, roughly half the emeralds would have been green if it hadn't been a law that emeralds are green. So if, by improbable coincidence, the first observed emeralds where in that 'would have been green' half we are also in an unlucky world (but there is a very small chance of this happening as well).

7 A Counterfactual Model of Inductive Knowledge

According to the prior knowledge theory, in the good cases, we are in a position to rule out some hypotheses about the emeralds colours prior to investigation. If $e_1...e_{100}$ are the first emeralds we will observe and they suffice for us to be able to acquire inductive knowledge, we should, at minimum, be able to rule out the possibility that $e_1...e_{100}$ will all be green by fluke. However it is *prima facie* unlikely that this is the *only* possibility we could rule out prior to investigation. For if it were only the fluke sequence of $e_1...e_{100}$ being green that I could rule out, and had I instead observed emeralds $e_2...e_{101}$, I would not be in a position to know that it's a law that emeralds are green after making those observations. It should strike us as too much of a coincidence that the exact possibility I in fact need to rule out to make my inductive inference is available to me, and nothing more. At the other extreme, we have the view that observing any 100 emeralds would have been sufficient to knowledgeably conclude that it's a law that emeralds are green. This view runs afoul of our third skeptical puzzle.

[48] This informal idea can be made precise probabilistically. This involves another vexed topic in the philosophy of conditionals concerning the probabilities of conditionals, which I am sidestepping for now. I treat the matter more fully, in a way that is consistent with the above reasoning, in Bacon (2015).

What we would like is a model of what we are in a position to rule out prior to investigation. An adequate model ought to *predict* that, in most worlds, inductive knowledge is possible from a relatively small number of observations, and ought also to predict that when the counterfactuals don't cooperate, inductive knowledge isn't possible. In the rest of the paper we will explore a simple model of inductive knowledge that has these features, drawing on our prior discussion of counterfactuals. Let us suppose that there is a particular sequence of colours — the distribution of colours the emeralds would have had if it hadn't been a law that emeralds are green — which we shall call s. Our model is a possible worlds model:[49]

> THE SIMPLE MODEL: The worlds open to you prior to investigation are (i) the lawlike worlds, (ii) the random worlds that differ from s about the colours of emeralds in at most k places.

Roughly, if we can determine that the number of differences between s and actuality exceeds some margin, k, we can knowledgeably conclude that the emeralds aren't random. (We are in a position to know that the differences between actuality and the distribution that would have been the case, had it not been a law that emeralds are green, does not exceed k.)

The counterfactual appealed to in the above is subject to the same sorts of qualifications we imposed in section 5: it must in particular be the sort of counterfactual expressed when the safety-theoretic notion of nearness is salient. In what follows, I will focus on the prior knowledge theory. However, a similar model can also be applied to the bonus knowledge theory: it could be that we are able to make the non-logical inference to the conclusion that it's a law that emeralds are green only when we have made enough observations to determine that differences between actuality and s exceed the margin.

To get a rough intuition for why this model might be appropriate, recall our example involving the lost keys. In the bad case, I don't get to know that my keys are in the car after searching the table and the mantlepiece. This is because, had they been in the house rather than the car, they would have been down the back of the sofa. But if I check the back of the sofa, and other similar places that the keys might have been had they been in the house, I presumably am in a position to know. The underlying thought is that I have to rule out the places where the keys *would (or might) have been, had they been in the house* before I get to know they're not in the house. Similarly, to know that it's a law that emeralds are green, I have to first rule out the sequences that would (or might) have obtained had it not been a law that emeralds are green.

The version of the theory I explore here assumes the outlook of conditional excluded middle discussed in the last section — that there is a unique world

[49] This means that ones total knowledge at a world w can be represented by a set of worlds, $K(w)$. A proposition modeled by a set of worlds, and a proposition p is known at a world w iff $K(w) \subseteq p$. Such models have some strong closure assumptions hardwired into them: if propositions $p_1, ..., p_n$ are individually known at a world w then every consequence of them (i.e. any superset of $p_1 \cap ... \cap p_n$) is also known at w. We will continue to use the terminology of worlds being *open* or *ruled* for an agent at a world w, which just means that the world in question either is or isn't in $K(w)$ respectively.

that would have obtained if it hadn't been a law that emeralds are green. In order to know that it's a law that emeralds are green you must rule out the sequence that would have obtained, had it not been a law, and those sequences that are relevantly similar (which I've represented here by saying that they do not differ from it by more than k places).

A fairly trivial variant of this model, that doesn't rest on conditional excluded middle, would instead pick a *set* of worlds that represent the worlds that *might* have obtained if it hadn't been a law that emeralds are green.[50] Again, the results would be similar: on the assumption that the worlds that might have obtained are all similar — say, differ by no more than some fixed number of places — we get a variant of our theory. Many of the general sorts of results that we obtain for the theory we explore here will have obvious analogues in this variant theory; we won't pursue them in the interest of keeping our discussion short.

More formally, a model is specified by the following data:

1. A set of worlds, W, with a designated world $@ \in W$.

2. A function $|\cdot|$ that associates each $w \in W$ with an assignment of colours to $e_1 ... e_{1000}$, $|w|$. (So $|w|$ is a sequence whose elements are either *GREEN* or *BLUE* of length 1000).

3. A selection function $f : (P(W) \setminus \{\emptyset\}) \times W \to W$ such that (i) $f(A, w) \in A$ and (ii) if $w \in A$, $f(A, w) = w$.

4. A distinguished subset $L \subseteq W$ of *lawlike* worlds, subject to the constraint that if $w \in L$ then $|w|$ is a constant assignment (every emerald is green or every emerald is blue).

5. A function $K : W \to P(W)$ such that $w \in K(w)$ for all $w \in W$.

$K(w)$ represents the agents total knowledge at w. $|w|$ represents the sequence of colours that $e_1 ... e_{1000}$ have at w. f is the counterfactual selection function. For example, on a similarity theoretic analysis of counterfactuals, $f(A, x)$ may be understood as the closest world to x where A is true. In what follows we are mainly interested in $K(@)$ where $@$ is the actual world. We shall continue to write s for $|f(\neg L, @)|$. (On the variant model which is friendly to failures of CEM, $f(A, w)$ picks out a set of worlds: the set of worlds that might have obtained had A obtained at w.)

[50] The choice over whether to formulate our principles in terms of 'would' or 'might' counterfactuals does not really matter if we are assuming CEM. Given CEM, the distinction between 'would' and 'might' conditionals is destroyed, assuming that 'would' and 'might' counterfactuals are duals. That they are duals just means that $p \Diamond\!\!\rightarrow q = \neg(p \Box\!\!\rightarrow \neg q)$. If this holds then $(p \Diamond\!\!\rightarrow q) = \neg(p \Box\!\!\rightarrow \neg q)$ implies $p \Box\!\!\rightarrow q$ by CEM. Conversely, if p is counterfactually consistent (i.e. $\neg(p \Box\!\!\rightarrow \bot)$) then $p \Box\!\!\rightarrow q$ entails $\neg(p \Box\!\!\rightarrow \neg q) = (p \Diamond\!\!\rightarrow q)$. Thus the only time would and might counterfactuals come apart is when the antecedent is counterfactually inconsistent. On the other hand, given the anti-CEM picture it's important to use might-counterfactuals: there could be no sequence of colours that *would* have obtained if it hadn't been a law that emeralds are green, but many that might have obtained had this not been a law.

Our models are characterized by the following constraint, the latter of which is just part of our description of the set up:

> $K(x) = L \cup \{w \in (W \setminus L) \mid$ the number of differences between $|f(\neg L, x)|$ and $|w|$ does not exceed $k\}$.
>
> $@ \in L$ and $|@|$ maps every emerald to $GREEN$

A word of warning is in order. The above is supposed to model what one is *in a position* to know, not what agents typically in fact know. The latter is subject to all sorts of idiosyncrasies relating to the agent's psychology. What a person is in a position to know, by contrast, is more objective and is in principle available to anyone in the same situation who forms the relevant belief for the right sorts of reasons.

This model predicts the two things I argued it should predict. Firstly, that inductive knowledge is possible in most worlds. A sequence of colors is assigned to a lawlike world if it is the distribution of colors the emeralds would have had if it hadn't been a law that emeralds are green at that world.[51] Assume that these sequences are distributed evenly over lawlike worlds: that the same number of worlds get counterfactually assigned a given sequence for any sequence. In the model this means that the number of x such that $|f(\neg L, x)| = t$ is non-zero and the same as the number of x such that $|f(\neg L, x)| = t'$ for any two sequences t and t'.[52] Assume, moreover, that the margin k is significantly less than 100: let's say it's 10. Then most lawlike worlds will be worlds in which there are k non-green emeralds among $e_1 ... e_{100}$ according to the counterfactual distribution of colors assigned to that world. (I.e. most lawlike worlds x are such that the sequence $|f(\neg L, x)|$ contains 10 or more non-green emeralds.) Since blue and green are the only possible colors, the proportion of worlds where there are less than k green emeralds in s is one in $\sum_{i \leq k} \frac{100!}{i!(100-i)!}$ when W is finite. This is extremely small when $k = 10$ (if there are more possible colors that green and blue, it gets even smaller). It follows, then that in most worlds in which it's a law that emeralds are green, we can come to know this by observing $e_1 ... e_{100}$ to be green. For in most lawlike worlds, s will assign k or more non-green colours to $e_1 ... e_{100}$, so observing that in fact $e_1 ... e_{100}$ are all green will allow us to determine that the number of differences from s exceeds k. Secondly, our model predicts bad cases where the counterfactual facts are inhospitable to inductive knowledge. In the rare cases where $e_1 ... e_{100}$ are mostly green in the counterfactually selected world (i.e. there are no more than k non-green members), we will not be able to determine that the differences between s and actuality exceed k.

[51] Note that at worlds where it isn't a law, this sequence is just the distribution of colors over worlds.

[52] We could also make these quantity claims using chances; the overall thrust of the argument would be similar.

7.1 Symmetry

Although this model is simple and predictive, I do not claim that it is primitively compelling or obvious. In this section and the next I will derive the main features of the model from two other attractive principles, lending it some independent plausibility.

The first is a very general sort of symmetry principle concerning the emeralds. Say that a distribution of colours over $e_{i_1}...e_{i_n}$ is *consistent* with a complete sequence of colours for all emeralds if $e_{i_1}...e_{i_n}$ have that distribution of colours in the sequence. That is to say, if d is an assignment of colours to $e_{i_1}...e_{i_n}$, it is consistent with the assignment d^+, assigning colours to every emerald, iff d^+ is an extension of the assignment d to all emeralds. Here $i_1...i_n$ and $j_1...j_n$ are two sequences of n different indices:

Symmetry If there is a distribution of colours over $e_{i_1}...e_{i_n}$ that is consistent with exactly N epistemically open sequences, then there is a distribution of colours over $e_{j_1}...e_{j_n}$ that is consistent with exactly N epistemically open sequences.

A sequence, t, is epistemically open iff there is a $w \in K(@)$ such that $|w| = t$. When $n = 1$ the principle encodes a sense in which the emeralds are all on a par with each other: we don't have more information about certain emeralds than others prior to investigation. If it were false there would be hypotheses about the colour of emerald e_i that rule out more sequences of colours than any hypothesis about the colour of e_j. Similar thoughts hold when $n > 1$: any two collections of n emeralds are on a par, prior to learning anything about the emeralds.

However we must proceed with extreme caution, since we have already seen that some symmetry intuitions are unreliable. Most saliently: the principle Emerald Anonymity discussed in section 4.2. Emerald Anonymity is in fact a special case of the following more general principle that has a similar form to Symmetry:

> If the claim that $e_{i_1}...e_{i_n}$ are all green is consistent with exactly N epistemically open sequences, then the claim that $e_{j_1}...e_{j_n}$ are all green is consistent with exactly N epistemically open sequences.

Emerald Anonymity is effectively the special case where $N = 1$.[53] Emerald Anonymity, when $n = 1$ does not directly encode the symmetry of the emeralds, rather it encodes a symmetry about certain emerald-colour *pairings*. Namely, the hypothesis that e_i is green is just as informative as the hypothesis that e_j is green: learning either would allow one to rule out the same number of worlds. But if e_i *would* have been green, and if e_j *would not* have been green, had it

[53] For if there is exactly one epistemically open world consistent with $e_{i_1}...e_{i_n}$ being green, and since the actual lawlike world is open (since knowledge is factive), then we know enough to conclude that all emeralds are green by law. Thus the $N = 1$ version of the principle is Emerald Anonymity.

not been a law that emeralds are green, it's not at all obvious that we should treat these pairs symmetrically given the considerations of the last section.

The emerald-colour pairings are not symmetric, rather the emeralds alone should play symmetric roles, and this is exactly what Symmetry says in the $n = 1$ case. If *some* hypothesis about e_is color rules out a certain number of worlds then some hypothesis about e_js colour also rules out that number of worlds.

I find it very hard to see how Symmetry could fail. However one might have the attitude that, given the failure of Emerald Anonymity we shouldn't rely on any symmetry intuitions at all: once bitten, twice shy. I think this is too sweeping a moral to draw. We have provided concrete mechanisms by which symmetry can be broken between emerald-colour hypotheses, but we have no such mechanisms for breaking symmetries between the emeralds themselves. Moreover, we have a compelling diagnosis of why we are inclined to be misled by Emerald Anonymity: a related principle in the vicinity is true, namely Symmetry, which is easily conflated with Emerald Anonymity.

7.2 Counterfactual Independence

Just as our first principle is in the vicinity of Emerald Anonymity, the second principle is in the vicinity of the principle Epistemic Independence discussed in section 4.1.

Counterfactual Independence If it's epistemically open that $e_{i_1}...e_{i_n}$ have a certain distribution of colors by fluke, and $e_{i_{n+1}}$ would have been green (blue) had it not been a law that emeralds are green, then it's epistemically open that $e_{i_1}...e_{i_n}$ has that distribution and that $e_{i_{n+1}}$ be green (blue).

Here i_{n+1} is distinct from $i_1...i_n$, and as before the counterfactual is subject to the same qualifications made above.

In what follows t, t', t'' etc. range over complete assignments of colours to emeralds, d, d' etc. over partial assignments, and we'll say that t (or d) is epistemically open if there's a (non-lawlike) world w that's epistemically open where the emeralds have that assignment of colours (i.e. $t = |w|$). Say that t is *at least as close* to s (which, recall, is $|f(\neg L, @)|$) as another t' iff whenever t' agrees with s about an emerald, so does t. Counterfactual Independence implies that if t is epistemically open, and t' is at least as close to s as t, then t' is epistemically open. For suppose t is epistemically open, and that t' is as close or closer to s than t. Consider X, the union of the set of emeralds that t and s agree about with the set of emeralds that t' and s disagree about. And consider the distribution d on X determined by t: this distribution on X is epistemically open because t is. Let $e_1...e_r$ be the remaining emeralds not in X. Without loss of generality, suppose e_1 is blue in s. It follows that the distribution d on X, augmented with e_1 being blue is also epistemically possible, so there is an epistemically open world w'' that has this distribution on $X \cup \{e_1\}$ (a similar conclusion can be reached if e_1 is green in s). Since t' is closer to s than $|w''|$, we can repeat the above reasoning for $e_2...e_r$ and conclude that the distribution

that agrees with t' on X and agrees with t' on $e_1...e_r$ is epistemically open. Since X and $e_1...e_r$ accounts for all emeralds this means that t' is possible.

7.3 Deriving the Model

Given these two principles we are in a position to derive the main features of our model. That is, these principles together entail that for some k, the epistemically open worlds prior to investigation consist in (i) some worlds in which it's a law that all emeralds are green (and some worlds where it's a law that all emeralds are blue), and (ii) some non-lawlike worlds that differ from $s = |f(\neg L, @)|$ about the colors of j emeralds for each j with $0 \leq j \leq k$.

Let's say that the *rank* of a sequence is the number of differences between it and s.[54] To show our claim it suffices to show that

(†) If t is epistemically open and t' has the same rank as t, then t' is epistemically open.

To see why this suffices, let k be the largest rank that an epistemically open non-lawlike world has. Any sequence t that differs from s by at most k places can be extended to a sequence t' that differs from s by *exactly* k places in such a way that t is closer to s than t' (this can be done just by making changes to t at places where t and s agree). It follows, by Counterfactual Independence, that t is possible if t' is. Finally, t' has the same rank as an epistemically open world (the world with the highest rank), so by (†), t' is epistemically open. This secures (ii). Since it's actually a law that emeralds are green, this is an epistemically open world. The parenthetical part of (i) — that it's epistemically open that emeralds are blue by law before making observations — does not follow, but is independently plausible given the way we have set the experiment up.

It suffices, then, to show (†). Suppose that $t \in K$ is epistemically open and that $r(t) = r(t')$ where these are the ranks of t and t' respectively. Now consider the emeralds $e_{i_1}...e_{i_n}$ on which t and s agree. If d is the distribution that t and s determine over $e_{i_1}...e_{i_n}$, then there are exactly $2^{r(t)}$ open sequences consistent with d on $e_{i_1}...e_{i_n}$, not counting the actual sequence of all greens where the law is in place. Reason: t is one such distribution, and moreover (by Counterfactual Independence) every distribution closer to s than t is also epistemically possible, of which there are $2^{r(t)}$. These are moreover all and only the distributions that are consistent with d: any open sequence that agrees with t on $e_{i_1}...e_{i_n}$ can only differ from t by agreeing with s at places where t does not, since the only emeralds not among $e_{i_1}...e_{i_n}$ are emeralds which t and s disagree about. Now let $e_{j_1}...e_{j_n}$ be the n emeralds on which t' and s agree (there are also n of these, since t' and t have the same rank — i.e. number of differences from s). By Symmetry it follows that there is a distribution d' over $e_{j_1}...e_{j_n}$ that is consistent with exactly $2^{r(t)} = 2^{r(t')}$ sequences, not including the actual sequence. It follows that every distribution that extends d' is epistemically open,

[54] We assume that ranks are assigned only to worlds where there are no laws about the colours of emeralds — we shall not assign a rank to the actual world.

since there are $2^{r(t')}$ sequences over the remaining emeralds in total (whether epistemically open or not). In particular the distribution t'' that extends d' by using t's assignment of colours to the remaining emeralds is epistemically open. Finally, since t' is closer to s than t'' it follows that if t'' is open, so is t'.

This shows that the worlds epistemically open prior to investigation have the right sort of structure. However there is a degenerate case to consider, in which s not open, and the only open worlds are lawlike. If any non-lawlike world is open, however, s is open because by Counterfactual Independence, if anything closer to s than an open world is also open, and so in particular if there's any non-lawlike open world s is open.

8 Conclusion

According to the prior knowledge theory, one must be in a position to know, prior to observation, that certain hypotheses about the colors of emeralds do not obtain. If this is correct, then there is a responsibility to say something a bit more systematic: if some hypotheses are ruled out and others aren't, it's natural to want to know which features determine which hypotheses are ruled out, and to explain how these features could be epistemically relevant. The challenge is significant because there is a tight symmetry between the possible distributions of colors over emeralds in non-law like worlds, analogous to the symmetries between possibilities in the lottery paradox.[55] It is hard to imagine what sorts of features could distinguish some sequences from others whilst also having a plausible claim to governing what one is in a position to know.

The counterfactual theory explored in the last section breaks these symmetries by drawing on the familiar connections between knowledge and counterfactuals. Of course, when some property separates things related by a symmetry, it often feels like an arbitrary line has been drawn. The counterfactual facts, on our picture, pick out a special distribution of colors over emeralds.[56] One might argue that this only shifts the bump in the rug: counterfactuals might explain why the distributions are not all epistemically on a par, but we have not provided a theory of why certain distributions are promoted over others by the counterfactual facts. Even so, we have reduced one sort of arbitrariness, to another more familiar sort. And perhaps it is unreasonable to ask for more from a theory of inductive knowledge: the question of what makes counterfactuals true is vexed, and is not something that necessarily needs to be settled before we can avail ourselves of a counterfactual theory of inductive knowledge.

References

Bacon, A. (2014). Giving your knowledge half a chance. *Philosophical Studies* (2), 1–25.

[55] See Hawthorne (2003) §1.2 for a discussion.

[56] Or perhaps, a special collection of distributions, if we are following the more general semantics for counterfactuals in which there can be more than one closest world.

Bacon, A. (2015). Stalnaker's thesis in context. *Review of Symbolic Logic 8*(1), 131–163.

Dorr, C., J. Goodman, and J. Hawthorne (2014). Knowing against the odds. *Philosophical Studies 170*(2), 277–287.

Fraassen, B. C. V. (1984). Belief and the will. *Journal of Philosophy 81*(5), 235–256.

Hawthorne, J. (2003). *Knowledge and Lotteries*. Oxford University Press.

Hawthorne, J. and M. Lasonen-Aarnio (2009). Knowledge and objective chance. In P. Greenough and D. Pritchard (Eds.), *Williamson on Knowledge*, pp. 92–108. Oxford University Press.

Huemer, M. (2001). *Skepticism and the Veil of Perception*. Lanham: Rowman and Littlefield.

Lasonen-Aarnio, M. (2010). Unreasonable knowledge. *Philosophical Perspectives 24*(1), 1–21.

Nozick, R. (1981). *Philosophical Explanations*. Harvard University Press.

Pryor, J. (2000). The skeptic and the dogmatist. *Noûs 34*(4), 517–549.

Sosa, E. (1999). How to defeat opposition to moore. *Philosophical Perspectives* (13), 141–154.

Wedgwood, R. (2013). A priori bootstrapping. In A. Casullo and J. Thurow (Eds.), *The A Priori In Philosophy*, pp. 226–246. Oxford University Press.

White, R. (2006). Problems for dogmatism. *Philosophical Studies 131*(3), 525–557.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford University Press.

Williamson, T. (2009). Probability and danger. In *Amherst Lecture in Philosophy*, pp. 1–35.