Lewis's Empiricism*

Wolfgang Schwarz
19 June 2021

Abstract. Like many forms of classical empiricism, Lewis's account of knowledge, mind, and language leaves no room for enquiry into non-contingent matters. According to Lewis, there is only one necessarily true proposition, and it is vacuously known by everyone. What, then, are we doing when we do metaphysics, which often seems to deal with non-contingent questions? Lewis never gave a satisfactory answer, or even acknowledged the problem. I explore some options. Can we understand the relevant parts of metaphysics as dealing with contingent questions about our concepts? Can we understand them in terms of a hyperintensional conception of "explicit" knowledge? Both options have some promise. Ultimately, however, I don't think they succeed at squaring Lewis's empiricist epistemology with his metaphysical realism.

1 Introduction

Great philosophers have often defended an epistemology that does not fit their own philosophical practice. Lewis is no exception. His epistemology, along with his philosophy of mind and language, leaves no room for substantive inquiry into non-contingent matters. Yet, for much of his career, Lewis seemed to be engaged in just that kind of inquiry, especially in his metaphysical works. What did he think he was doing? And whatever he thought, how can we make sense of Lewisian metaphysics if we find ourselves attracted, as I do, to his epistemology?

I will begin by reviewing Lewis's epistemology. In many ways, it is a classical empiricist epistemology, of a kind that made other philosophers skeptical or hostile towards metaphysics. Lewis, of course, was neither skeptical nor hostile towards metaphysics. I will explore some attempts at resolving this tension, but I have no easy answer to offer.

2 Lewis's epistemology

Within epistemology, Lewis is best known for his contextualist analysis of knowledge ([Lewis 1979b], [Lewis 1996]). But contextualism is only a small and somewhat tangential

* A precursor of this chapter was presented at the Workshop *David Lewis and His Place in the History of Analytic Philosophy* in Manchester. Thanks to everyone who gave me feedback. Special thanks to Mahrad Almotahari, Helen Beebee, Ralf Busse, Wolfgang Freitag, and Thomas Sattig.

part of his epistemological viewpoint. The bigger picture is often mentioned as an aside, while discussing other topics or addressing specific puzzles – see especially [Lewis 1986a: 27-40], [Lewis 1979a], [Lewis 1994b: 308-324], [Lewis 1983b: 49-55], [Lewis 1983a], and [Lewis 1974]. Many elements of this bigger picture were not Lewis's invention but adopted from others, including Carnap, Hintikka, Stalnaker, and Jeffrey (see, e.g., [Carnap 1962], [Hintikka 1962], [Stalnaker 1984], [Jeffrey 1992]). Let me recapitulate the main ideas.

We start with the modal conception of knowledge and belief. When we're ignorant of something, we don't know whether the world is one way rather than another. Some ways the world might be are compatible with our information. Others are not. This suggests that we may capture an agent's knowledge by specifying a class of possible worlds: all the worlds that are compatible with their information. Likewise for belief. An agent's belief state represents the world as being a certain way; we can represent its content by the class of worlds that are that way.

What does it take for an agent to stand in the belief relation to a class of possible worlds? No mysterious grasp of possibilia is required. It is also not required that the agent somehow accepts a sentence in a public or private language, which in turn expresses the relevant class of worlds. For Lewis, the possible-worlds concept of belief is rather defined by its functional role: an agent is belief-related to a certain class of worlds iff they are in a state that plays the right functional role.

This role has several components. One connects belief states to desire states and behaviour. Roughly: agents are disposed to act in a way that would bring them closer to satisfying their desires if the world were as they believe it to be. Another part of the belief role connects belief states to perceptual experience. Roughly: belief states tend to change through experience by incorporating new information about the perceived environment. In addition to these input and output conditions, Lewis posits a range of general eligibility constraints on what an agent may believe and desire.

These ideas can be spelled out in different ways. Lewis preferred a broadly Bayesian approach, in which the binary notion of belief is replaced by a graded notion of credence – a probability measure over possible worlds. The output connection between belief, desire, and behaviour is then spelled out by decision theory: a joint state of belief and desire normally disposes an agent to choose acts that maximize expected utility relative to their credence and utility functions. The input connection between belief and perceptual experience becomes a form of Bayesian updating. In its simplest version, we assume that each perceptual experience is associated with a class of worlds; when an agent has the experience, all her credence is shifted to that class in a way that preserves credence ratios within the class. To ensure that the resulting belief state is reasonable in light of the agent's evidence, further eligibility constraints are needed. Among other things, Lewis suggests that (absent unusual evidence) little credence should be given to scenarios in which perceptual experiences are highly unreliable or in which the observed part of the

universe is radically different from the unobserved.

For Lewis, these constraints, like the input and output rules, are both normative and constitutive. They are normative insofar as they describe a rational ideal of which real people often fall short. But they are also constitutive because they (or something much like them) implicitly define the concepts of belief and desire: an agent's physical state is correctly interpreted as such-and-such a state of belief and desire just in case the interpretation makes the agent come out closest to the rational ideal. In practice, no interpretation will have perfect fit, and there may be several interpretations that fit equally well. The agent's belief state will then be indeterminate between the candidates with best fit. (Indeterminacy also arises because the ideal itself is not fully determinate.)

Lewis assumes that the central psychological concepts in terms of which we interpret one another are belief and desire. The concept of knowledge plays a different and less important role. It serves as a "messy short-cut" [Lewis 1996: 440] to convey information about an agent's evidence. By evidence, Lewis says, he means experiences and (quasi-)memories. Strictly speaking, any world in which we have our current experiences and (quasi-)memories is compatible with our evidence. This includes worlds where we are brains in a vat. As a result, it is difficult to succinctly and accurately convey information about someone's evidence. The concept of knowledge helps by allowing us to "ignore" some parts of logical space: we may truly say that an agent knows a proposition as long as that proposition is true at all non-ignored worlds compatible with the agent's evidence.

Which worlds may be ignored depends on conversational context. If a skeptical possibility is made salient, it can't be properly ignored. This, Lewis suggests, explains why skeptical arguments appear so persuasive. As I mentioned above, I doubt that Lewis would have regarded this suggestion as central to his epistemology. Serious epistemology does not involve the messy shortcut of knowledge. Its main question is what credence one should give to which possibilities in light of which evidence – where no possibilities are ignored.

Bayesian models nicely account for the holism of evidential support. Whether a given experience raises or lowers an agent's credence in a given hypothesis usually depends on the agent's background assumptions. Lewis even argues that we should treat belief states as holistic units, without assuming that they consist of separate beliefs in distinct propositions. Traditional questions about the structure of justification, about a web-like or building-like justification relation connecting different beliefs, may therefore seem out of place. Nonetheless, one can find a clear affinity between Lewis's account and a kind of empiricist foundationalism.

Perceptual experiences (and quasi-memory) have a foundational status. Lewis, in fact, comes close to saying that they are known infallibly and with absolute certainty. This is an immediate consequence of the account described in [Lewis 1996], whereby everyone knows what experiences they have, no matter which possibilities are ignored.

In the standard Bayesian model, our evidence similarly has maximal credence: rational agents never doubt their perceptual input. Jeffrey [1965: ch.11] famously argued against this assumption, which he called "hardcore empiricism". Lewis had some sympathy for Jeffrey's complaint, but argued that the assumption is plausible for ideal agents (see [Lewis 1986b: 62f.]).

Our perceptual foundation supports the rest of our belief state through Lewis's substantive (non-formal) rationality principles. In standard Bayesian models, these are encoded in an agent's "ultimate priors" – their credence function before taking into account any evidence. Such a credence function, Lewis assumes, should give comparatively little weight to scenarios in which we are Boltzmann brains or deceived by evil demons. Most of its weight should go to well-behaved worlds in which our senses are reliable, in which other people's testimony is generally trustworthy, and in which physical processes conform to simple and systematic regularities.

Unlike some classical empiricists, Lewis allows us to extrapolate from present experiences not just to hypotheses about further experiences. He does not hold that every meaningful question can be translated into a question about experience. He has no reservations about positing unobservable entities. He takes scientific theories at face value and trusts the scientific method. (In particular, he is confident that the methods of physics can uncover the "perfectly natural properties" that set apart genuine from spurious regularities.) True hardcore empiricists may well frown upon these aspects of his epistemology. What, they might argue, could justify the assumption that science gives us access to unobservable parts of reality? Similar complaints could be made about Lewis's other rationality principles. What justifies the assumption that our senses are reliable, or that the unobserved world resembles the observed?

Since these assumptions are encoded in an agent's ultimate priors, it is clear that they are not meant to be justified by any relevant evidence. Lewis also admits that they are not supported by non-circular arguments. Skeptical scenarios in which our senses are unreliable or in which the methods of science lead us astray cannot be shown to be incoherent or unlikely without already presupposing that they are. In that sense, these prior assumptions might also be regarded as foundational. If true, they constitute an inconclusive and defeasible kind of a priori knowledge.

We also have conclusive a priori knowledge. But there is only one example, setting aside technical problems in measure theory: in Lewis's possible-worlds models, the class of all worlds is known and believed by everyone, irrespective and independent of their evidence, and no matter what is ignored; it always has credence 1.

Like many forms of classical empiricism, Lewis's epistemology does not allow for substantive knowledge of non-contingent facts. If a proposition is true at all worlds, then we automatically know it. Not because we have tremendous rational insight, but because there really isn't anything interesting to know here. All genuine knowledge is

contingent knowledge. Accordingly, gaining knowledge is always a matter of receiving new information from the senses. There is no provision for *reasoning* to change our knowledge or rational credence.

That's where Lewis's epistemology appears to clash with his philosophical practice.

3 The problem of metaphysical omniscience

When Lewis does philosophy, he often appears to be interested in non-contingent questions. He studies how to analyse causation, whether there are universals or tropes, whether other possible worlds are as real and concrete as the actual world, and so on. Lewis treated these questions as substantive and seemed to think that philosophy could find the answer.

He also explained how they should be approached (see, e.g., [Lewis 1983c: x-xi], [Lewis 1986a: 133-135], [Lewis 1973: 88f.], as well as [Nolan 2015]). We should, he says, look for an account of the relevant topic that is simple, systematic, and conservative, meaning that it should not deviate too far from previous opinion. Philosophical arguments may change the credibility of a hypothesis, but even at the end of inquiry, there will rarely be a decisive winner. Opposing views can always be upheld. "Once the menu of well-worked out theories is before us, philosophy is a matter of opinion." [Lewis 1983c: xi].

Lewis's empiricist epistemology would seem to predict a very different attitude. Any unambiguous, well-defined, and non-contingent hypothesis expresses either the empty set or the set of all worlds. Why should the choice between such hypotheses turn on the balancing of theoretical virtues (simplicity, systematicity, conservatism)? How could it be a matter of opinion?

More strikingly perhaps, Lewis often professes ignorance of non-contingent matters. In [Lewis 1986a], he says that he doesn't know whether there are duplicate worlds, how large a possible spacetime can be, or whether positive and negative charge are co-instantiated at some world. Lewis does not appear to be troubled by this kind of ignorance: "why should I think that I ought to be able to make up my mind on every question about possible worlds, when it seems clear that I may have no way whatever of finding out the answers to other questions about noncontingent matters – for instance, about the infinite cardinals?" ([Lewis 1973: 89]).

In Chapter 4 of *Plurality* ([Lewis 1986a: 109–115]), Lewis talks about our knowledge of modality and maths. He insists that any acceptable epistemology must account for our knowledge of maths and must do so without "giving mathematics some devious semantics". Perceptual experience, he suggests, is needed to locate ourselves within the space of possibilities but not to discover what there is from an objective perspective. Mathematical and metaphysical inquiry is concerned with this other question. He does not explain how

we are supposed to square these ideas with the modal account of knowledge defended in Chapter 2 of the same book.

What's odd about this tension – between Lewis's empiricism and his rationalism, if you like – is that it does not arise in obscure and distant parts of his philosophy. Lewis wrote a lot on mind and language, and he wrote a lot on metaphysics. Much of what he wrote on mind and language appears to clash with what he did in metaphysics. This is hardly a problem he could have overlooked.

In what follows, I will try to figure out what Lewis might have thought about the problem. I will also try to figure out what *I* should think about it. As I said, I find the empiricist epistemology attractive. Ignorance is lack of information. When we lack information, we don't know which possible way things might have been is actually the case. The information we lack is contingent information about the world around us. There is no other kind of information to lack. The idea that reality has further, non-contingent aspects to which we gain access by reasoning strikes me as mysterious and unnecessary. But then what should we say about apparently non-contingent domains of inquiry?

Let's explore a few options.

We could, of course, hold on to the empiricist epistemology and conclude that any apparent inquiry into non-contingent matters, such as Lewisian metaphysics, is "nothing but sophistry and illusion". That is as simple as it is unappealing. Let's set it aside. I will also set aside the opposite strategy, to declare the empiricist epistemology radically mistaken or incomplete – although I suspect this might have been Lewis's own preferred resolution.

I'm interested in less radical and less obvious solutions. For example, I have assumed that the kind of metaphysical inquiry that Lewis was engaged in is an attempt to answer non-contingent questions. It is worth double-checking this assumption.

Indeed, you might argue that the apparent tension simply arises from an equivocation. We can agree that metaphysical questions about universals or possible worlds are *metaphysically* non-contingent. But couldn't they still be *epistemically* contingent? If we don't assume that the "possible worlds" over which credence, belief, and knowledge are defined are metaphysically possible, we can allow for substantive knowledge and ignorance about metaphysics, without significantly changing the picture I have described in the previous section.

Or so you might think. But things are not that simple. There is by now an extensive literature on adding metaphysically impossible worlds to possible-worlds accounts of knowledge and belief – see, for example, [Nolan 2013], [Jago 2014], [Bjerring and Schwarz 2016], [Elliott 2019]. It is not hard to define a suitable concept of worlds. What is less clear is whether the resulting account yields a useful model of metaphysical (or mathematical) ignorance and inquiry, and to what extent this model would be continuous with the model from the previous section. Among other things, we would arguably need

a new account of what it is to stand in the belief relation to a class of worlds. And we would need a new account of how a credence measure over the extended class of worlds may rationally change in response to philosophical considerations.

Lewis, at any rate, always resisted introducing a distinction between (deep) epistemic and metaphysical possibility. He had only one space of worlds. Much of his metaphysics does not vary within that space. There aren't Lewisian worlds with Armstrongian universals, others with primitive laws, and still others with neither. There aren't Lewisian worlds where modal realism is true and others where it is false. (It is not even clear what that would mean.) Lewis argued that his one space of worlds is a "philosopher's paradise" because it can be used to analyse important philosophical concepts, including the concepts of knowledge, belief, and credence. He was a firm believer in what Chalmers [2006] calls the "golden triangle", linking meaning, reason, and modality.

At the very least, then, appealing to impossible worlds in this context would involve a substantial departure from Lewis's views. I will not directly explore the other issues it would raise, except to note that much of what I will cover in sections 5 and 6 below could be recast in terms of impossible worlds.

4 Reinterpretation

Introducing metaphysically impossible worlds is one way of making metaphysical inquiry look like empirical inquiry. Another way is to suggest that although metaphysical questions are often non-contingent, when we are trying to answer these questions, we are really trying to answer contingent questions. Let me explain.

Modal accounts of knowledge and belief imply that these attitudes are closed under logical consequence. If P is true at all "epistemically accessible" worlds, and P entails Q, then Q is also true at all epistemically accessible worlds. As a special case, logically necessary propositions are automatically known since they are entailed by everything and true at all worlds. This "problem of logical omniscience" is often regarded as a serious flaw. Real knowledge and real belief, it is assumed, are not closed under logical consequence. Real people don't know all logical truths, and they don't know the consequences of everything they know. If that were correct, we would have good reason to revise Lewis's empiricist epistemology, no matter what we think about metaphysics.

But is it correct? Lewis (like Stalnaker, see [Stalnaker 1984], [Stalnaker 1991], [Stalnaker 1999]), was not convinced. While he sometimes speaks of logical omniscience as an idealisation (e.g., [Lewis 1983b: 275]), he also defended the idea that real knowledge ([Lewis 1996: 441f.]) and belief ([Lewis 1986a: 32-36]) are closed under logical consequence (or at least, closed under consequence within each resolution of an indeterminate/fragmented belief state).

To be sure, one can easily think of cases in which we are inclined to say that someone

fails to know a complicated logical truth, or in which someone fails to know a consequence of something they know. But even granting that these judgments are true, they only bear on possible-worlds models if we assume some connection between ordinary attitude reports and attitudes understood in terms of possible worlds. Here, Lewis calls for caution.

It is tempting to assume that, on possible-worlds accounts, a statement of the form 'S knows that P' is true iff the embedded sentence P (or some proposition it expresses) is true at all worlds epistemically accessible for the subject denoted by S. Since, for example, 'there are infinitely many primes' is true at all worlds, it would follow that 'S knows that there are infinitely many primes' is true for any S. But clearly many people do not know that there are infinitely many primes.

Lewis rejects this connection between attitude reports and attitudes characterised in terms of possible worlds. The real connection, he says, is "complicated and multifarious" [Lewis 1986a: 34] (see also [Lewis 1979a], [Lewis 1981], [Lewis 1994b]).

For example, Lewis suggests that what is reported by 'Oscar knows that water is wet' is, roughly, that Oscar stands in a certain relation to water in which he (or rather his epistemic counterparts) stands to something wet at all his epistemically accessible worlds. (This move has become popular in formal semantics, following [Percus and Sauerland 2003].) One might likewise suggest that 'Oscar doesn't know that 3847 is prime' reports that there is a way in which the number 3847 is known to Oscar – perhaps as the number denoted by the numeral '3847' – in which a non-prime number is known to Oscar at some of his epistemically accessible worlds (compare [Cresswell and von Stechow 1982], approvingly cited in [Lewis 1986a: 35, fn.26]).

Relatedly, Lewis suggests that ordinary attitude reports sometimes have quotational or meta-linguistic interpretations. When someone "doesn't know that a fortnight is two weeks", arguably the real object of their ignorance is of a linguistic nature: they don't know that the word 'fortnight' means *two weeks*.

Once we break the simple link between attitudes and ordinary attitude reports, it is no longer obvious that our attitudes are not closed under logical consequence or that we are ignorant of non-contingent matters. Apparent ignorance of a necessary truth – that 3847 is prime, or that a fortnight is two weeks – may really be ignorance of contingent linguistic facts.

This is what I had in mind when I said that we might try to understand non-contingent metaphysical inquiry as inquiry into contingent questions. The real object of metaphysical inquiry would not be non-contingent hypotheses about causation, universals, or the nature of possible worlds, but contingent surrogate propositions, perhaps about words or concepts.

This interpretation looks more attractive for some parts of Lewis's metaphysics than for others. The parts for which it looks more attractive are those in which he offers an

analysis – of causation, laws of nature, chance, dispositions, knowledge, belief, conventions, values, and so on.

Lewis usually puts his analyses in the material mode: 'C causes E iff so-and-so'. On the present proposal, the true object of this proposal would be a contingent hypothesis about our words or linguistic dispositions. Perhaps it is the hypothesis that we are disposed to apply the word 'cause' to a pair C, E iff so-and-so (or something more complicated along similar lines).

To be clear, the idea is not that this contingent hypothesis is semantically expressed by the original statement, 'C causes E iff so-and-so'. Rather, the idea is that asserting that statement somehow expresses a belief in the contingent proposition (perhaps through a Stalnakerian process of "metalinguistic diagonalisation").

The proposed reinterpretation arguably makes sense of our practice of discovering and responding to counterexamples. Imagine at some point we believed that knowledge is justified true belief; then we came across Gettier cases, where we found ourselves disposed to judge that certain cases of justified true belief are not cases of knowledge. If the object of our original belief was a non-contingent hypothesis about the nature of knowledge, it is not entirely clear why contingent information about our dispositions (to assent to certain strings of symbols) should be relevant. Arguably, we made a surprising discovery not about the non-contingent nature of knowledge, but about our linguistic dispositions, and thereby about other people's dispositions, and thereby about the meaning of the word 'knowledge'.

The reinterpretation strategy also makes superficial sense of Lewis's claims about philosophical methodology.

For one, it is understandable why conservatism plays a role. If an analysis deviates widely from our prior opinions, it is unlikely to capture our dispositions for applying the relevant terms. As Lewis says in [Lewis 1997] on the analysis of colour terms: "It won't do to say that colours do not exist; or that we are unable to detect them; or that they never are properties of material things". If you say any of these things, it is doubtful whether you are offering a correct account of our linguistic dispositions.

It is also easy to understand why metaphysical inquiry should follow similar standards as empirical inquiry. On the present interpretation, metaphysical inquiry *is* empirical inquiry. It's a special type of empirical inquiry that can be largely carried out from the armchair.

We may also explain why there are no knockdown arguments and why contrary views can always be upheld. This is the well-known underdetermination of theory by observation. Faced with an apparent counterexample, you can always claim that our intuitions in this case are mistaken – say, that our concept of knowledge really does apply to Gettier cases, but that we get confused when we think about these cases. (We know that we are not perfect in applying our concepts.)

There is even some direct evidence that Lewis endorsed the reinterpretation strategy for his philosophical analyses. In a 1999 letter to Mary Kate McGowan (letter 110 in [Beebee and Fisher 2020: 218f.]), he explains that conceptual analysis is an investigation into our classificatory dispositions. His goal, he says, is to uncover his classificatory dispositions; he is putting forward "fallible hypotheses about an independent (mental) reality that I can't examine directly".

Nonetheless, I don't think the reinterpretation strategy fully succeeds at reconciling Lewis's empiricist epistemology with his rationalist metaphysics.

Remember the problem. Lewis's official epistemology leaves no room for ignorance of, or inquiry into, non-contingent matters. Gaining knowledge is always a matter of ruling out previously open possibilities. But then how can we make sense of metaphysical inquiry, which often appears to deal with non-contingent matters? According to the reinterpretation strategy, this appearance is misleading: when we debate apparently non-contingent metaphysical hypotheses, the true object of our inquiry are contingent surrogate propositions about our linguistic dispositions. (Compare: when two people argue over whether a fortnight is 12 or 14 days, the true object of their disagreement is arguably whether 'fortnight' means 12 or 14 days, which is contingent.)

One obvious problem for this response is that it still leaves no room for the process of a priori inquiry. Perhaps some progress in metaphysics and maths can be understood as the empirical discovery of linguistic facts. But much of it seems to be a matter of reasoning, not observation.

To illustrate the problem, consider a simple arithmetical task – say, to check whether 29 is prime. To do this, you might check whether 29 is divisible by 2, then whether it's divisible by 3, then whether it's divisible by 5, and then note that $7^2 > 29$, so there's no point trying other primes. All these steps are easy. For example, you see that 29 is not divisible by 2 by noticing that the last digit, '9', is not even. What happens at these steps? Are you really gaining contingent information at each of them?

You might argue that whenever you figure out a mathematical truth S, then you really are discovering contingent information about the world – for example, that you are disposed to assent to S when given sufficient time to think etc. But why didn't you already have this information? Couldn't you all along rule out worlds in which the last digit of '29' is not '9' (if there are such worlds)? Couldn't you rule out worlds in which you are not disposed to say that a number that ends in '9' isn't divisible by 2? You may not have consciously thought of these matters before starting the computation, but it is hard to believe that you lacked the relevant information. (It is also hard to believe that you had the information, but that it was distributed over different fragments of your belief system. Among other things, since computations can get arbitrarily long, there would have to be an unbounded number of fragments.)

On closer inspection, the reinterpretation strategy also struggles to make sense of

the theoretical virtues that govern metaphysical inquiry. If the true subject matter is our dispositions to use a word or concept, we should look for simple and systematic theories *of our dispositions*. This is not tantamount to looking for simple and systematic analyses of the relevant concepts. For example, a systematic hypothesis about why we are disposed to make certain judgments might appeal to universal cognitive mechanisms, such as our tendency to use spatial concepts when reasoning about non-spatial issues. Yet considerations like these are alien to Lewis's metaphysics.

An especially puzzling virtue, from the perspective of the reinterpretation strategy, is ontological parsimony. Lewis thinks an analysis that posits fewer types of things is, all else equal, better (see, e.g., [Lewis and Lewis 1970]). But why should ontological parsimony be relevant if the object of inquiry is our linguistic dispositions? Note that the ontological parsimony of an analysis does not make the surrogate proposition about our dispositions any more parsimonious.

Recall also that Lewis regarded some hypotheses in maths and metaphysics as not just unknown but unknowable. That's odd if the relevant hypotheses are really hypotheses about our linguistic dispositions. It is tempting to think that there's a recursive enumeration of all our linguistic dispositions, since they are generated by an effectively finite computational engine.

Relatedly, and most simply, there is more to metaphysics (and maths) than conceptual analysis. Take modal realism. According to Lewis, there are at least beth-2 electrons, all of which are spatiotemporally isolated from one another. That doesn't look like an analysis of anything. What is the contingent surrogate proposition supposed to be?

In sum, the reinterpretation strategy may have some promise for some aspects of Lewisian metaphysics, but on its own, it does not succeed at resolving our puzzle.

5 Explicit knowledge

It's not hard to understand the project Lewis is engaged in when he defends metaphysical doctrines like modal realism. He starts with the miscellany of statements that we are inclined to accept, in everyday life or in more theoretical contexts. Some of these appear to quantify over possible worlds. As a metaphysician (and Quine student), Lewis wants to know what we should make of these statements. Can apparent quantification over possible worlds be paraphrased away, perhaps in a fictionalist manner? Lewis says no: only with serious costs. But it is useful talk, so we also don't want to give it up. We should therefore accept that there really are possible worlds. We must then ask what kinds of things they are. Are they *sui generis* entities that cannot be described in any other way? That's "magical ersatzism", and Lewis argues that it, too, has serious costs. Lewis suggests that our quantification over worlds is best understood as quantification

over maximal spatiotemporally related mereological fusions. We should enrich our total theory by adding these commitments.

In general, much of Lewis's metaphysics tries to systematise certain aspects of our total theory. He wants to see which terms can be defined by others, and how the things we talk about in one fragment of our theory relate to the things we talk about in other fragments. The goal is to find a simple, elegant, perspicuous successor to our starting point.

What's being systematised in this project are not Lewisian propositions – classes of worlds. The starting point are sentences. The task is similar to the familiar mathematical project of axiomatising a given (mathematical or physical) theory, of finding a minimal set of primitive terms and axioms from which the entire theory can be generated by means of definitions and logical inference – except that we also allow for small revisions if that helps the axiomatisation.

It's plausible that this project involves language. But it's not plausible that the project boils down to finding out contingent facts *about* language. As I mentioned above, whatever the relevant contingent facts might be, it is hard to see why we didn't already know them from the start.

When Lewis talks about non-contingent knowledge, he often suggests that this is a matter of accepting statements or theories (e.g., [Lewis 1986a: 113ff.]). This is in sharp contrast to his normal epistemology and philosophy of mind, where he insists that knowledge and belief should not be understood in terms of a putative relation (of "acceptance" or otherwise) to sentences.

Lewis, therefore, appears to work with two concepts of knowledge and belief, one sentential and one non-sentential. How are these concepts related?

In [Lewis 1986a: 27ff.], he alludes to a distinction between "explicit" and "implicit" knowledge and belief. He does not explain the distinction. He merely points out that his modal account is not restricted to explicit attitudes. Some such distinction has become popular in epistemic logic (see e.g. [Fagin and Halpern 1987]), but most of the relevant work was written after Lewis's comments, so he was probably not referring to any particular account in that literature. Let me try to spell out what he might have had in mind.

Physical agents must store information in a physical format. The format in which information is stored makes a difference to its use. Some representational formats make it easy to apply a piece of knowledge to a given task, others require lengthy computations. If a lengthy computation is needed, and you don't have much time, you can't perform the task, even though you have the information. You have the information, but in the wrong form.

When asked if 29 (or 2529) is prime, I have all the information needed to answer the question. I don't need to observe my dispositions or inspect Plato's heaven to answer the

question. But I need to do some computations to convert the stored information into a 'yes' or 'no' answer.

Possible-world models of knowledge represent what information an agent has, abstracting away from the format in which the information is stored. For ideal agents, this makes no difference. But for non-ideal agents, it does. If a non-ideal agent has the information P stored in the wrong way, they may sometimes not be disposed to act in a way that would bring them closer to satisfying their desires if P were true.

We might say that an agent "explicitly knows that P" if they have stored the information expressed by P in the indicated format, as the sentence P. But that would commit us to the contentious assumption that the agent stores information in the form of English sentences. It might be better to say that an agent "explicitly knows that P" if they know (the proposition expressed by) P in a way that goes along with a relatively effortless disposition to affirm a sentence that is synonymous with P (and that the agent understands).

For non-contingent truths, implicit knowledge is vacuous; explicit knowledge here reduces to something like a disposition to affirm a suitable sentence (which one understands). This is not far from Lewis's suggestion, briefly floated in [Lewis 1986a: 113], that someone knows that 2+2=4 iff they "fully understand and accept the statement". Explicit belief would be the same, except that the target sentence needn't be true.

With these concepts in play, we might hope to accommodate a priori inquiry without compromising the core ideas of Lewis's empiricist epistemology. On the picture that emerges, mere reasoning does not provide us with new information, but it might provide us with new explicit knowledge.

The purpose of a priori metaphysics would not be to figure out whether the world is one way or another. Rather, it would be to find a new linguistic representation of what we already (implicitly) know. The goal is to convert the sentences we accept into a succinct, systematic (linguistic) description of reality, with a small ideology and a qualitatively sparse ontology.

Why should we be interested in this project? There are pragmatic reasons. Simplicity helps clear thinking. It's hard to reason in an intensional language with 257 primitives; it's easy to make mistakes in such a language and to miss connections between the various primitives.

But is there more to it? Are simple and elegant metaphysical theories more likely to be true?

6 Best systems

Metaphysical realism, let's say, is the view that metaphysics deals with substantive, irreducible, and often non-contingent questions about fundamental reality. Whether there

are mereological fusions, whether there is a fundamental relation of causation, whether other worlds are as real and concrete as the actual world – for the realist, the answers turn on the mind-independent metaphysical structure of reality. The aim of metaphysics is to discover that structure.

Throughout his work, Lewis appears to endorse this realist perspective. But I fear that it does not mesh with his empiricist epistemology, even after we have made room for hyperintensional "explicit" attitudes along the lines I suggested in the previous section. When we convert one linguistic representation into another, how could we thereby gain access to a mind-independent aspect of reality?

Metaphysical realism appears to require at least three further assumptions. First, it requires not just a hyperintensional conception of knowledge, but also a hyperintensional conception of facts. We need to distinguish, say, modal realism and mereological universalism, even if both are true (or both are false) at all possible worlds. Second, we need an explanation of how our words manage to latch on to such facts. The words 'there are universals', for example, must be associated with a particular hyperintensional fact, so that accepting the words somehow amounts to believing that the relevant fact obtains. (Lewis's convention-based account of how words get their meaning arguably doesn't extend to this case.) Finally, we need an explanation of how accepting the words can amount to *knowledge* of the relevant facts. (At one point in [Lewis 1986a: 113], Lewis suggests that hyperintensional knowledge might require nothing more than accepting a statement that is true. But that doesn't seem right. Could I really come to *know* that there are universals or large cardinals merely by accepting a corresponding statement, for no good reason?)

None of these assumptions were part of the extended empiricist account that I outlined in the previous section. And I don't think they could easily be added. The account instead suggests a very different perspective on non-contingent metaphysics, on which the project is simply to systematise certain aspects of our total theory, and that's all there is.

A comparison with physics may be useful. Physics, let's assume, tries to discover the fundamental laws of nature. But how should we understand these laws? Many have thought that they constitute an irreducible nomic aspect of reality (see, e.g., [Maudlin 2007]). Lewis disagreed. He did not believe that reality has a fundamental nomic dimension. According to Lewis, what physicists are trying to discover when they are trying to discover the laws is an elegant systematisation of occurrent events – of the "Humean mosaic", or more generally, of the pattern of instantiation of fundamental properties and relations in the world.

Perhaps metaphysics can be seen as continuous with this enterprise. Physics attempts to systematise the contingent patterns in our world. Physicists are not interested in non-contingent aspects of such a systematisation. If two physical theories make all the

same claims about particles and fields etc., but one of them quantifies over sets and the other doesn't, or one quantifies over universals and the other doesn't, they are regarded as merely different formulations of the same theory. But metaphysicians care about the difference. Metaphysics, one might suggest, is concerned with non-contingent aspects of our total theory – as well as some contingent aspects on which scientists can't be trusted. Just as the aim of physics is not to uncover a hidden nomic layer of reality beyond the Humean mosaic, the aim of metaphysics, on this account, is not to uncover a hidden layer of metaphysical structure beyond the Humean mosaic.

I have defined metaphysical realism so that it implies commitment to non-contingent metaphysical facts (about the existence of universals or the nature of possible worlds, for example) over and above facts about the distribution of fundamental properties and relations in the world around us. We might want to allow for weaker forms of metaphysical realism. Lewis regarded himself as a realist about physical laws, even though he did not believe in irreducible law facts. According to Lewis, there really are physical laws. All it takes for something to be a law is that it figures in the best systematisation of the Humean mosaic. We might similarly say that something is a "metaphysical law" if it is a non-contingent part of the best systematisation. The task of metaphysics would be to find these laws.

Lewis noticed that his best-systems account of laws seems to give rise to a kind of anti-realism or relativism, which he saw as the account's most serious drawback: "when we ask where the standards of simplicity and strength and balance come from, the answer may seem to be that they come from us" [Lewis 1994a: 232]. In response, he argued that our standards aren't arbitrary: there are objective facts about comparative simplicity and strength. But he conceded that "*if* disagreeing rival systems were running neck-and-neck" – a possibility he regarded as far-fetched – "th[e]n lawhood might be a psychological matter, and that would be very peculiar." [Lewis 1994a: 233]

In the case of metaphysics, the possibility of rival systems running neck-and-neck does not look far-fetched at all. Lewis himself suggested that there may be an approximate tie between positing universals and positing a primitive resemblance relation, in terms of whatever theoretical virtues govern metaphysical inquiry. Since these virtues are somewhat vague, one person might prefer the theory with universals while another prefers primitive resemblance. On the best-systems account of metaphysics, neither of them would be objectively right or wrong. There would be no fact of the matter.

Lewis would no doubt have regarded this as a drawback (compare [Lewis 1983c: xi]). But it might be a price worth paying, if it frees us from the burden of adding mysterious epicycles to an otherwise attractive picture of mind, language, and the world.

There are other costs, however. In Lewis's best-systems account of physics, the statements that are part of candidate systems are not statements about a problematic nomic dimension of reality. Instead, they are assumed to make straightforward and

unproblematic claims about the contingent distribution of fundamental properties and relations. ('All Fs are Gs', as the toy example goes.) There is no difficulty understanding what it would take for such a statement to be true. Metaphysical statements, by contrast, frequently seem to describe a problematic metaphysical dimension of reality. Consider 'there are universals', or 'any two things have a mereological fusion'. If we don't think reality has a hyperintensional, mind-independent metaphysical structure, how should we interpret these statements, as they might occur in a candidate system? What would it take for them to be true?

Perhaps truth here amounts to no more than being part of the best system. This is similar to what Lewis says about chance. According to Lewis, what makes it true that an event has chance x is that the best system says that it has chance x. One might nonetheless find it odd to apply this to a large chunk of metaphysics – especially since metaphysical hypotheses often aren't expressed in terms of special concepts (analogous to chance) that one might take to be defined with recourse to a best system. Lewis's conjecture that there are at least beth-2 spatiotemporally isolated electrons, for example, looks like a straightforward conjecture not about what is part of our best system, but about the total number of electrons.

The best-systems picture here leads to a Carnapian sort of doublethink. Suppose our best theory says that there are universals, and non-contingently so. We should then talk as if we think that reality has that particular metaphysical structure, even though we should also recognize that, in some sense, reality has no extra metaphysical structure at all.

All in all, I doubt that Lewis would have liked this picture. I suspect he would have opted for a more radical departure from his official theories of mind and language. He would have preferred an account that makes room for substantive hyperintensional facts which are somehow expressed by metaphysical and mathematical statements and which can be the object of hyperintensional knowledge. He nowhere offers even a hint of how that account might go.

I am not optimistic about the prospects for such an account. I am also not convinced that we need it. Granted, we often think and talk as if reality had a hyperintensional structure to which we can gain access by reasoning. But perhaps this is an aspect of our total theory that we should be willing to revise.

References

Helen Beebee and Anthony Fisher (Eds.) [2020]: *Philosophical Letters of David K. Lewis: Volume 1 Causation, Modality, Ontology*. New York: Oxford University Press, first edition. ISBN 978-0-19-885545-3

Jens Christian Bjerring and Wolfgang Schwarz [2016]: "Granularity Problems". *The Philosophical Quarterly*, 67(266): 22–37

Rudolf Carnap [1962]: "Logical Foundations of Probability"

David Chalmers [2006]: "The Foundations of Two-Dimensional Semantics". In Manuel Garcia-Carpintero and Josep Macia (Eds.) *Two-Dimensional Semantics,* Oxford: Oxford University Press, 55–140

Maxwell J. Cresswell and Arnim von Stechow [1982]: "De Re Belief Generalized". *Linguistics and Philosophy*, 5(4): 503–535

Edward Elliott [2019]: "Impossible Worlds and Partial Belief". *Synthese*, 196(8): 3433–3458

Ronald Fagin and Joseph Y. Halpern [1987]: "Belief, Awareness, and Limited Reasoning". *Artificial intelligence*, 34(1): 39–76

Jaakko Hintikka [1962]: *Knowledge and Belief: An Introduction to the Logic of the Two Notions.* Ithaca: Cornell University Press

Mark Jago [2014]: *The Impossible.* Oxford: Oxford University Press

Richard Jeffrey [1965]: *The Logic of Decision.* New York: McGraw-Hill

— [1992]: *Probability and the Art of Judgment.* Cambridge: Cambridge University Press

David Lewis [1973]: "Counterfactuals and Comparative Possibility". *Journal of Philosophical Logic*, 2: 418–446

— [1974]: "Radical Interpretation". *Synthese*, 23: 331–344

— [1979a]: "Attitudes *De Dicto* and *De Se*". *The Philosophical Review*, 88: 513–543

— [1979b]: "Scorekeeping in a Language Game". *Journal of Philosophical Logic*, 8: 339–359

— [1981]: "What Puzzling Pierre Does Not Believe". *Australasian Journal of Philosophy*, 59: 283–289

— [1983a]: "Individuation by Acquaintance and by Stipulation". *The Philosophical Review*, 92: 3–32

— [1983b]: "New Work for a Theory of Universals". *Australasian Journal of Philosophy*, 61: 343–377

— [1983c]: *Philosophical Papers I.* New York, Oxford: Oxford University Press

— [1986a]: *On the Plurality of Worlds.* Malden (Mass.): Blackwell

— [1986b]: "Probabilities of Conditionals and Conditional Probabilities II". *The Philosophical Review*, 95: 581–589

— [1994a]: "Humean Supervenience Debugged". *Mind*, 103: 473–490

— [1994b]: "Reduction of Mind"

— [1996]: "Elusive Knowledge". *Australasian Journal of Philosophy*, 74: 549–567

— [1997]: "Naming the Colours". *Australasian Journal of Philosophy*, 75: 325–342

David Lewis and Stephanie Lewis [1970]: "Holes". *Australasian Journal of Philosophy*, 48: 206–212

Tim Maudlin [2007]: *The Metaphysics within Physics*. Oxford; New York: Oxford University Press. ISBN 978-0-19-921821-9 0-19-921821-8

Daniel Nolan [2015]: "Lewis's Philosophical Method". In B. Loewer and J. Schaffer (Eds.) *A Companion to Lewis,* Wiley-Blackwell, 25–39

Daniel P. Nolan [2013]: "Impossible Worlds". *Philosophy Compass*, 8(4): 360–372

Orin Percus and Uli Sauerland [2003]: "On the LFs of Attitude Reports". In *Proceedings of Sinn Und Bedeutung,* vol 7. ISBN 2629-6055, 228–242

Robert Stalnaker [1984]: *Inquiry*. Cambridge (Mass.): MIT Press

— [1991]: "The Problem of Logical Omniscience, I". *Synthese*, 89(3): 425–440

— [1999]: "The Problem of Logical Omniscience II". In *Context and Content,* 255–273