Wolfgang Schwarz

Modal Metaphysics and Conceptual Metaphysics

1

Metaphysics is, at least in part, a quest for the fundamental elements of reality, for the basic truths that determine all the rest. Some hold that this basis is purely physical; others say that truths about conciousness or moral rightness are not determined by the physical facts. Some say that the physical truths can be further reduced to the distribution of local physical qualities: once God—to use the familiar image—has settled which qualities are located where, not just at the beginning of time, but for all times, there is nothing left to do; others maintain that he (or she) would still have to settle facts about causation and the laws of nature.

These debates can be understood in different ways. On one reading, they deal with *conceptual* relations: the question is whether—and if so, how—truths about consciousness or causation are conceptually entailed by more fundamental truths. Call this *conceptual metaphysics*. On another reading, what is at issue are *modal* relations: whether truths about consciousness or causation are strictly implied, or necessitated, by the other truths; that is, whether every possible world that agrees with our world about such-and-such facts also agrees with it about consciousness or causation. Call this *modal metaphysics*.

That modal and conceptual entailment come apart was demonstrated by Saul Kripke (1980): given that water is necessarily H_2O, every possible world where there is H_2O on Mars is also a world where there is water on Mars; so "there is H_2O on Mars" necessitates "there is water on Mars." But there is no conceptual or otherwise a priori entailment between the two.[1]

Consequently, philosophers have endorsed claims in modal metaphysics while rejecting the corresponding conceptual claims. In particular, some

[1] I move freely between "a priori" and "conceptual" in this paper, dodging the great old question whether some a priori truths might be synthetic.

physicalists in the philosophy of mind hold that the physical truths necessitate the mental truths even though there is no conceptual connection between the two. (This is called "type-B" physicalism in David Chalmers's [2002] taxonomy.)

2

Such views have recently come under attack. In *From Metaphysics to Ethics*, Frank Jackson argues that

> physicalists, *qua* holders of a metaphysical [i.e. modal] view, are committed to the logical thesis of the a priori deducibility of the psychological way things are from the physical way things are. (Jackson 1998, 57)

David Lewis generalizes:

> *all* of us are committed to the a priori deducibility of the manifest way things are from the fundamental way things are. (Lewis 2002, 96)

Jackson and Lewis do not deny that there are a posteriori necessities. Rather, they argue that a proper analysis of a posteriori necessities undermines the apparent distinction between conceptual and modal metaphysics.

Consider the a posteriori necessity "if there is H_2O on Mars, then there is water on Mars." This is a posteriori: we need empirical information, that water is H_2O, to figure out that it is true. How can we get that information? By taking a water sample and analysing its chemical composition. This will not conclusively establish that water is H_2O, however, as we might have been unlucky in our choice of the sample; perhaps all the watery stuff in our rivers and lakes is XYZ except for a few traces of H_2O that show up under rare conditions, and our sample happened to be of this type. To rule that out, we could take more samples. What matters is this: if we knew that *all* the watery stuff in our rivers and lakes is H_2O (or that almost all of it is predominately H_2O), we could conclusively infer that water is H_2O. And from that we could infer that if there is H_2O on Mars, then there is water on Mars. That is, a posteriori necessities like this follow *a priori* from ordinary, contingent information about the actual world.

Here is another way of making the point. Let us mean by "the proposition expressed" by a sentence S something that determines for each possible world w whether "at w, S" is true or false. Then we need empirical information to find out what proposition is expressed by "water is H_2O": a

proposition true at every possible world. The information we need, however, is not mysterious modal information, to be acquired by verneoscopic insight into other possible worlds. Rather, it is ordinary information about our world and our surroundings—about what the stuff in our rivers and lakes consists of.

The so-called 'two-dimensional' analysis of a posteriori necessities says that this is true for all cases:

> **2D**: for every a posteriori necessity S there is some ordinary, contingent truth C about the present context such that $C \supset S$ is a priori.[2]

This is sometimes presented as part of a general two-dimensional semantics for English, but the general theory will play no role in what follows. One can accept 2D without accepting any of these more general theories.

Jackson (1998, 81-83) now argues as follows. Suppose P necessitates Q, but the entailment is a posteriori. Then by 2D, the entailment can always be turned a priori if P is enriched with further contingent information about the present context. Hence if we begin with some P that already entails *every* contingent truth, there is no such further information to add. In such a case, therefore, the entailment cannot be a posteriori. That is,

> **ModCon**: whenever some truths necessitate all truths, then they conceptually (or a priori) entail all truths.

3

ModCon is false. Here is a counterexample.

> **Triv**: All things are exactly as they actually are.

Triv sounds trivial, but it is not true at every possible world. At many worlds, things are different from how they actually are—that is, from how they are here in this world. At some worlds, the mountains are higher, the stars brighter, the cockatoos quiet. Not so in the actual world. In the actual world, all things are exactly as they actually are. So Triv is true, and trivial, but contingent.

What do other worlds look like where Triv is true? They are perfect duplicates of the actual world. For whenever a world differs from our world in

[2] "ordinary" is meant to rule out truths like "water is H_2O iff the meter bar is one meter long" that would trivialise the claim; see Lewis 2002.

any way, things over there are not exactly as they are here. Assuming that different worlds are never perfect duplicates of one another—if exactly the same things are true in two worlds, why are they *two* possibilities and not *one*?—, we can simplify: Triv is true at the actual world and nowhere else.

It follows that Triv necessitates every truth. For any truth S and any world w, if Triv is true at w, then S is true at w. (Suppose, for reductio, that S is false at w. Then things in w are not exactly as they are in our world. So Triv isn't true at w. Contradiction.)

By ModCon, we can infer that Triv conceptually entails every truth. But evidently it does not. Triv doesn't tell us who will win the next elections, how many species of penguins there are, or whether 'tis nobler in the mind to suffer the slings and arrows of outrageous fortune, or to take arms against a sea of troubles. Indeed, it doesn't tell us anything. That is why we don't need to know the first thing about the world to figure out that Triv is true.

As ModCon is false, something must be wrong with Jackson's argument. The problem is the move from "P entails every truth" to "we need no further information about the present context to find out that $P \supset Q$." If P only *necessitates* every truth, we may well need such further information. Triv, for example, necessitates everything, and in this sense entails every truth. But it doesn't tell us anything about the present context, and therefore doesn't tell us what propositions our sentences express.

Here is the relevant passage in *From Metaphysics to Ethics*:

> if physicalism is true, all the information needed to yield the propositions being expressed about what the actual world is like in various physical sentences can be given in physical terms, for the actual context is givable in physical terms according to physicalism. Therefore, physicalism is committed to the in principle a priori deducibility of the psychological from the physical. (Jackson 1998, 93)

This isn't true. Understood as a claim about modal entailment, physicalism only says that all facts about the actual context are *necessitated* by what can be given in physical terms. It doesn't follow that all these facts are *a priori deducible* from the physical description.

4

So modal metaphysics and conceptual metaphysics really do come apart. Let's have a closer look at how this can happen.

Return to Triv. In second-order two-dimensional modal logic, this can be expressed without any non-logical terms:

Triv: $\forall x \forall F(Fx \leftrightarrow \text{actually } Fx)$.

The logic of "actually" guarantees that this is true in (the designated world of) every model. In this sense, Triv is a *pure logical truth*: a logical truth expressed in purely logical vocabulary. (By contrast, "either it rains or it doesn't rain" is an impure logical truth, containing non-logical terms.)

As we've seen, all truths are necessitated by Triv. Therefore, *all truths are necessitated by pure logical truths.*

Should we conclude that the fundamental nature of our world is purely *logical*; that once the logical truths are settled, everything else is fixed as well? Could a lazy creator have created our world by merely fixing the logical truths—by declaring "the world is to be just as it actually is!," and then going to rest? That seems wrong.

The trouble is caused by the "actually" operator that allows us to express very rich propositions in a very intransparent way, by refering back to how things are in the actual world when evaluating a sentence at other possible worlds. "Mount Everest is just as high as it actually is" is true at a world w iff Mount Everest is 8848 meters high at w. In this sense, the sentence says that Mount Everest is 8848 meters high. But it does so in such an intransparent way that it is completely useless for telling somebody the height of Mount Everest, or for setting its height in the first place when creating the world.

The same problem arises with rigidified descriptions. Suppose some kind of Cartesian dualism is true and *being conscious* is a non-physical property whose instantiation causes certain events N in the pineal gland. Suppose these events are never caused in any other way. Let T be the expression "the property whose instantiation is the actual cause of N." T rigidly denotes *being conscious*: at any world, the property whose instantiation is the *actual* cause of N is the property of being conscious. Hence we can use T to determine, in seemingly innocent physical terms, that someone is conscious: "someone has the property whose instantiation is the actual cause of N." If this counts as a physical truth, the physical truths end up necessitating truths about consciousness. Should we conclude that Cartesian dualism is a form of *physicalism*?

No. Instead, we should insist that serious metaphysics must not rely on back-reference to the actual world: if you say that all truths are necessitated by some truths A (the physical truths, or whatever), what proposition is expressed by A should be conceptually independent of how things are in the actual world. A should make sense even when used by the lazy creator in creating our world.

5

If back-reference to the actual world is allowed, physicalism becomes trivial. As we've seen, even Cartesian dualism will qualify. Indeed, assuming that logical truths count as physical truths, Triv guarantees that everything is necessitated by the physical truths, no matter what the world is like—no matter if it is riddled with ghosts and gods and immaterial souls.

Philosophers usually characterize physicalism as something like the claim that any world that agrees with our world in all physical respects also agrees with it in all mental respects (see Jackson 1994, §2; Chalmers 1996, 38-41). This is OK, as long as we are careful about what counts as a 'physical respect': the fact that Triv or "someone has T" is true should not count as a physical feature of our world.

The most common device for modal back-reference in English are proper names. Let "@" rigidly denote the actual world. Then instead of Triv, we can say "everything is just as in @." Again, this necessitates every truth. Likewise, on the Cartesian view above, we could introduce a new name for what actually causes N and use it to describe what is the case at our world. Such expressions, too, must be banned from a properly physical basis.

With these restrictions in place, it turns out that panpsychism and Russellian monism ("type-F monism" in Chalmers [2002]) no longer count as physicalist positions. According to these views, physical terms such as, say, "charge -1," rigidly pick out essentially mental properties. That is, for something in another possible world to have charge -1, it must have that mental property; it doesn't matter if it has a property that plays the physical role charge -1 plays in our world. "charge -1" thus works very much like our name for what actually normally causes N: when you say that something at w has charge -1, you *seem* to characterize it in physical terms, but since "charge -1" back-refers to whatever plays the charge -1 role in our world—which happens to be an essentially mental property—, what you say determines that something in w has this mental property, and leaves it open whether it has any property playing (in w) the role of charge -1.

It has always been a curious artefact of the usual definitions that panpsychism and Russellian monism, with their fundamental mental properties, qualify as physicalism. Once we realize that they do so only by exploiting modal back-reference, we have good reason to exclude them from the physicalist family.

6

What about type-B physicalism, the target of Jackson's argument? Does

it, too, secure its status as a physicalist position only by means of the back-reference trick that makes physicalism trivial?

If the two-dimensional analysis of a posteriori necessity is correct, the answer is "yes." Recall 2D: every a posteriori necessity is a priori entailed by ordinary, contingent truths about the actual world and our place in it. So if the implication $P \supset Q$ from the physical to the mental is necessary a posteriori, then its truth in other worlds conceptually depends on how things are in the actual world. This is a sure sign of modal back-reference: if a non-modal sentence does not contain back-refering devices, then whether or not it is true at a world w conceptually depends only on how things are *at* w, not also on how they are at the actual world.

Some type-B physicalists reject 2D. They maintain that some necessities are not only a posteriori, but also impossible to settle by any ordinary information about our world and our surroundings. Chalmers (2002) calls such necessities *strong necessities*.

From a formal point of view, it is easy to see how there could be strong necessities. We only need to decouple the space of possible worlds from any links to epistemic or conceptual matters. Suppose, for instance, that there are only 7 possible worlds, all of which contain penguins. Then the existence of penguins is a strong necessity: no ordinary investigations in the actual world, and no amount of conceptual analysis, could reveal to us that penguins necessarily exist, and yet they do.

The main problem with this view is that it ignores the constraints that come with the name "possible worlds." It is like saying that for all we know, there might be only 7 natural numbers, all of which are prime. This is nonsense: if some things don't at least satisfy the Peano Axioms, they don't deserve the name "natural numbers." Likewise, there may well be some space of world-like entities with only 7 members; but as these things won't do any of the jobs associated with possible worlds (see Lewis [1986, ch. 1] for a start), they don't deserve the name "possible worlds."

This might need a more lengthy defense than I can give here. Let me make another point instead. Suppose the doctrine of strong necessities is true and unrestricted modal discourse is about a realm of world-like entities which is not in any way tied to consistency, conceivability and what can be established by empirical investigations in the actual world. Then I fail to see why we as philosophers should pay attention to it. Suppose physicists discover that our universe is split into several spatiotemporally isolated parts. Would we care much about what is the case in the other parts? Would we deem it of great metaphysical importance whether all of them have penguins? I think not. But on the doctrine of strong necessities, possible worlds are exactly like this. In other words, I want to suggest that modal space is metaphysically relevant only because of its close ties to matters conceptual and epistemic.

7

I've used physicalism as an example, but what I said applies across the board. Take any claim of modal metaphysics: that all truths are necessitated by Φ-truths. It faces a trilemma. Either (1) the claim essentially exploits back-reference to the actual world, which robs it of its metaphysical significance. (If this is the game, simple logical truths suffice to determine everything.) Or (2) the claim relies on the doctrine of strong necessities, which also robs it of its metaphysical significance (as well as making it dependent on a doctrine that's arguably nonsense). Or (3) the claim coincides with the corresponding claim of conceptual metaphysics: that all truths are conceptually entailed by the Φ-truths.

The upshot is that Jackson was right after all: serious metaphysics is conceptual metaphysics.

It is worth noting that many branches of metaphysics never took the turn to the modal in the first place. For the metaphysics of non-contingent matters such as mathematics and modality, it is clear why: any necessary truth is trivially necessitated by everything, so the modal project is pointless here. Even with contingent facts, the modal dependence is often taken for granted, and the debate is largely about *how* the truths in question are made true by the more fundamental truths. A classic example is David and Stephanie Lewis's (1970, 1996) prototype of a metaphysical dispute, on the nature of holes: both parties agree that truths about holes are necessitated by truths about non-holes; this isn't the end of the dispute—it's the starting point (see especially Lewis & Lewis 1996, 184f.). Or think of the discussion over whether facts about objective chance and the laws of nature are determined by the distribution of fundamental Humean qualities. Claiming that this holds simply as a matter of brute necessity (or securing it with the back-reference trick) would not count as playing by the rules. If you say the dependence holds, you have to say *how* the Humean distribution determines the laws; you have to give an analysis of laws that makes the entailment analytic.

It is mainly in philosophy of mind (and some parts of meta-ethics) that Kripke's discoveries have been taken to show that metaphysics and conceptual analysis are radically different things. It is time to revert that. Kripke didn't break the link between metaphysics and conceptual analysis. If anything, he broke the link between metaphysics and modality.

REFERENCES

Chalmers, D. (1996). *The Conscious Mind*. New York: Oxford University Press.

Chalmers, D. (2002). Consciousness and its place in nature, *Philosophy of Mind. Classical and Contemporary Readings*, ed. by D. Chalmers, 247-272.

Jackson, F. (1994). Armchair metaphysics, *Philosophy in Mind*, ed. by J. O'Leary-Hawthorne & M. Michael, 23-42. Dordrecht: Kluwer.

Jackson, F. (1998). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Clarendon Press.

Kripke, S.A. (1980). *Naming and Necessity*. Oxford: Blackwell.

Lewis, D. (1986). *On the Plurality of Worlds*. Malden, MA: Blackwell.

Lewis, D. (2002). Tharp's third theorem. *Analysis* 62, 95-97.

Lewis, D. & Lewis, S. (1970). Holes. *Australasian Journal of Philosophy* 48, 206-212.

Lewis, D. & Lewis, S. (1996). Casati and Varzi on holes. *Philosophical Review* 105, 77-79. Page numbers refer to the reprint in Lewis's *Papers in Metaphysics and Epistemology*, 1999.