

Category Mistakes and Ordinary Language

Martin Gustafsson

1.

Gilbert Ryle starts off his essay “Categories” by noting that category-propositions—assertions that terms belong to certain categories—are characteristically *philosophical*.¹ He does not mean simply that such propositions are always philosopher’s propositions. The converse is also true, says Ryle: philosopher’s propositions are always category propositions (C 189). If this is right, the present paper deals not merely with a special and delimited philosophical problem but with the nature of philosophy itself.

Ryle’s claim is a stipulative provocation, of course. However, he is undoubtedly right that category distinctions, and mistakes due to failures to properly observe such distinctions, have been central concerns of philosophy since ancient times. Traditionally, such concerns have been seen as part of metaphysics. And I agree with Ryle that much traditional metaphysics—metaphysics in its pre-Quinean, pre-Kripkean, pre-Lewisian form—is basically a matter of exploring category distinctions and making clear wherein they consist (C 203).

To the extent that the exploration of category distinctions constitutes the centerpiece of metaphysics, it would seem that a version of that endeavor can survive the Kantian attack on so-called “dogmatic” metaphysics.² Dogmatic metaphysics proceeds from a notion of reality as separate from what the dogmatic metaphysician sees as our merely subjective or

¹ Gilbert Ryle, “Categories,” *Proceedings of the Aristotelian Society* 38:1 (1938), p. 189; henceforth C, followed by page number.

² Kant prefaces the second edition of the *Critique of Pure Reason* with a discussion of the need for a critique of the dogma of metaphysics. See Immanuel Kant, preface to *Critique of Pure Reason*, trans. Paul Guyer and Allen W. Wood (Cambridge: Cambridge University Press, 1998), Bxxx–Bxxxvii. In the final chapter of the *Prolegomena* he spells out the difference between “metaphysics as science”—a science whose “entire stock of *a priori* concepts” has been set forth by “a critique of reason itself” according to their sources in sensibility, understanding, and reason—and the “dogmatic chatter” with which one had to be satisfied before one had “tasted of critique” (*Prolegomena to Any Future Metaphysics*, ed. and trans. Gary Hatfield (Cambridge: Cambridge University Press, 1997), 119-120).

conventional means of representation. Thus, the dogmatic metaphysician takes it to be both possible and appropriate to investigate the world “from sideways on,”³ to use John McDowell’s suggestive phrase—that is, to occupy a standpoint from which one somehow gains more direct access to the world than when one relies on the conceptual resources provided by language or thought. By contrast, the sort of non-dogmatic or “critical” metaphysics that Kant does not seem to want to undermine proceeds from the notion that the metaphysical order of reality—the only reality we can have any positive conception of at all—is *the same* as the logical order of thought. According to the critical metaphysician, this order is an order that thought and reality must share if there is going to be anything like *thinking about what is in the world* at all. Since we *can* think about what is in the world, there must be such a shared logical order. If the critical metaphysician is right—then metaphysics can be meaningfully pursued only from within language or thought, and that pursuit will *be* an investigation of that shared logical order. And, since category distinctions are traditionally supposed to be distinctions that are present in both reality and in thought or language, it seems that the exploration of such distinctions may well take an appropriately critical shape and thus constitute a sort of metaphysical inquiry that philosophers in a broadly Kantian tradition would be perfectly willing to accept.⁴

Indeed, category distinctions are just as much a topic of logic as they are a topic of metaphysics. In early twentieth-century analytic philosophy, the logical treatment of categories took the shape of type-theoretical considerations, beginning with Russell’s and Whitehead’s heroic effort to save the logicist project from inconsistency.⁵ Wittgenstein’s discussion of what he calls “formal concepts” in the *Tractatus*⁶ is another manifestation of how important issues of this kind were to the founders of the analytic tradition, even if Wittgenstein thought that no theory of types could bolster or justify the categorial classifications apparently designated by formal concepts. In fact, according to Wittgenstein, categorial or formal distinctions cannot be designated or justified at all (TLP 4.126). Rather, such distinctions are exhibited in the employment of language. Thus, Wittgenstein holds that

³ John McDowell, *Mind and World* (Cambridge: Harvard University Press, 1994), p. 34.

⁴ See, for example, Sebastian Rödl, *Categories of the Temporal: An Inquiry into the Forms of the Finite Intellect* (Cambridge: Harvard University Press, 2012), p. 40.

⁵ See Alfred North Whitehead and Bertrand Russell, *Principia Mathematica* (Cambridge: Cambridge University Press, 1919).

⁶ Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, trans. D.F. Pears and B.F. McGuinness (London: Routledge and Kegan Paul, 1961), esp. 4.126–4.1273; henceforth TLP, followed by section number.

in a syntactically adequate notation, there will simply be no expressions for formal concepts; instead, there will be rules for the use of signs (TLP 3.343–4). Perhaps one might say that according to early Wittgenstein, the very idea that category distinctions can be designated and justified itself constitutes an especially tempting and harmful kind of category mistake. For early Wittgenstein, there can be no meaningful category-propositions—such “propositions” are misguided attempts to say what can only be shown in the meaningful use of words.

As traditionally conceived, category distinctions are in no way empirical or hypothetical. Rather, they are *a priori* conditions for the very possibility of meaningful thought and talk about reality. After all, we cannot investigate just ‘entities’ in general. For anything is a mere entity, so to speak; if I come to the world with no further principles of discrimination, I will not be able to distinguish any determinate subject of inquiry at all. Certainly, I might investigate the physical realm without already knowing about emeralds, asteroids, snowflakes, or quarks. But if I have no notion of what a *physical object* is, my investigation will not yet have found its subject matter—I will not be able to ask questions about physical objects *qua* physical objects, and thus I will have no clear idea of what I am looking for.

One way in which the special character of category distinctions manifests itself is in the fact that if the character of such distinctions is not underscored, it will be difficult to provide a satisfactory account of negation. Consider the following sentence:

(1) The number 2 is taller than 1,500 feet.

Now, imagine an interlocutor who said, “This sentence is false, and blatantly so; but it is in principle no different from other falsehoods, such as ‘The Empire State Building is taller than 1500 feet’ or ‘Stockholm is a larger city than Paris.’” This kind of view would entail that the negation of sentence (1) is true:

(2) It is not the case that the number 2 is taller than 1,500 feet.

But then, consider the following sentence:

(3) The number 2 is either exactly 1,500 feet tall or it is shorter.

On the face of it, (2) is equivalent to (3). However, it would seem pretty arbitrary to say that (1) is false and (3) is true! For it is just as weird to say of the number 2 that it is exactly 1,500 feet tall or shorter as it is to say of it that it is taller than 1,500 feet. Our imagined interlocutor may try to get out of this quandary by arguing that whereas statements (1) and (2) are contradictory, (1) and (3) are not genuine contradictories—but mere contraries that can be and in fact are simultaneously false—thus claiming, in effect, that saying of the number 2 that it is not taller than 1,500 feet is altogether different from saying that it is either exactly 1,500 feet tall or it is shorter. The problem with this maneuver is that this is not how negation works in cases where the subject *is* clearly of the right category. For example, take the sentence

(2*) It is not the case that the Empire State Building is taller than 1,500 feet.

This *is* equivalent to

(3*) The Empire State Building is either exactly 1,500 feet tall or it is shorter.

So, it seems that our interlocutor will have to distinguish between two kinds of negation, one of which is present when the length of number 2 is at issue, and the other of which is present when the height of the Empire State Building is at issue. But then this interlocutor owes us a clarification of the difference between these two senses of negation, together with an account of how we are to decide in a given case which sense of negation is relevant. And it is hard to see how she can provide such an account without invoking what are, in effect, category distinctions! So, her attempt to assimilate category mistakes to ordinary falsehoods does not seem to work. The need for category distinctions will not go away; thus, nothing seems gained by insisting that it is just blatantly false to say that the number 2 is taller than 1,500 feet. It seems better to say that such a sentence is nonsense, or at least that it lacks a truth-value, and that there just is not anything like negating what is expressed by such a construction, since such a construction expresses nothing true or false at all.⁷

⁷ See, for example, Leonard Goddard, “Nonsignificance,” *Australasian Journal of Philosophy* 48:1 (1970), pp. 10–6.

2.

I have tried to indicate, in an introductory, sketchy, and abstract fashion the significance that philosophers have traditionally taken category distinctions to have—why philosophers have found it important to identify such distinctions and to clarify mistakes that arise if we fail to observe them. I have also suggested that the exploration of category distinctions might naturally be taken to constitute a central part of a critical metaphysical endeavor—an investigation of the ultimate structure of reality *via* an investigation of the logical structure of thought or language.

I have not defended such a critical-metaphysical conception of what investigating category distinctions demonstrates. In fact, very many philosophers who acknowledge the philosophical importance of category distinctions have nonetheless rejected the idea that such distinctions are in any way distinctions “in the world.” According to these philosophers, a sound conception of category distinctions will have to distance itself not only from dogmatic metaphysics but from metaphysics altogether; and many have concluded that such distancing can be achieved only if we think of category distinctions as *merely* linguistic and not in any more substantive sense “real.” Carnap was such a philosopher. For him, questions of category should be reformulated as questions about which linguistic framework to choose; such questions are what he calls purely “pragmatic,” non-theoretical one—matters of convenience rather than truth.⁸ Ryle may seem to take a similarly anti-metaphysical line, arguing that “as only collocations of symbols can be asserted to be absurd or, consequently, denied to be absurd, it follows that category-propositions are semantic propositions” (C 206). However, he immediately adds that this does not “imply that [such propositions] can say nothing about the ‘nature of things,’” and then goes on to elaborate: “If a child’s perplexity why the equator can be crossed but not seen, or why the Cheshire Cat could not leave its grin behind it is perplexity about the ‘nature of things,’ then certain category-propositions will give the required information about the nature of things” (ibid.). At this point, Ryle may seem quite open to a sort of critical-metaphysical conception of what can be learnt from category-propositions. And one gets a similar impression from his central works, including *The Concept of Mind*.⁹ Clearly it is Ryle’s aim there to illuminate what skill, habit, intelligence, will, and so forth, *are*; he does not think of the significance of his investigations as somehow restricted to a merely linguistic realm.

⁸ See Rudolf Carnap, “Empiricism, Semantics, and Ontology,” *Revue internationale de philosophie* 4:11 (1950), p. 27.

⁹ See Gilbert Ryle, *The Concept of Mind* (London: Routledge, 2009), esp. pp. 8–12.

Austin even more explicitly resists the idea that his investigations into “what we should say when” are investigations of language *rather than* of extra-linguistic reality. As he famously notes, “When we examine . . . what words we should use in what situations, we are looking . . . not *merely* at words . . . but also at the realities we use the words to talk about: we are using a sharpened awareness of words to sharpen our perception of, though not as the final arbiter of, the phenomena.”¹⁰ But he does not explain further what this is supposed to mean, and in this paper I shall not discuss his viewpoint in any more detail.

What about Wittgenstein? *That* is a hard question to answer! To repeat, in the *Tractatus*, he believes that there are no meaningful category-propositions: such propositions are misguided attempts to say what can only be shown in the employment of signs (TLP 4.126). On the other hand, it is by no means clear that what is thus shown can be understood as belonging exclusively to language, *as opposed to* extra-linguistic reality: there is, I believe, little reason to think that early Wittgenstein’s notion of logic traffics in any such distinction between language and extra-linguistic reality. Logic, according to Wittgenstein, must “presuppose that names have meaning and elementary propositions sense” (TLP 6.124) and is therefore concerned with what is *not* arbitrary in our linguistic symbolism, as that symbolism is used to describe the world. There is a sharp difference between early Wittgenstein and Carnap on this point: for Wittgenstein, there is one world and one logic, and there is no such thing as investigating and evaluating their interconnection from sideways on. Early Wittgenstein stands closer to a broadly Kantian tradition of critical metaphysics than to the sort of anti-metaphysical conventionalism that Carnap can be said to represent. It is very hard to pinpoint exactly how close he is to this Kantian tradition or to what the points of connection amount. I shall have a bit more to say about this later, but there will be no space to give anything like a fully adequate account of early Wittgenstein’s viewpoint.

Later Wittgenstein abandons the sort of logical absolutism that is so central in the *Tractatus*. In the *Philosophical Investigations*, it is emphasized over and over again that there is not just one logic but an open-ended and dynamic variety of grammars that are intricately interwoven with the practical activities of real-life human beings.¹¹ However, it is far from clear that this conception involves a notion of category distinctions as *merely* linguistic and not in any more substantive sense “real.” Certainly, grammar is not dictated by some

¹⁰ J.L. Austin, *Philosophical Papers*, 3rd ed., ed. J.O. Urmson and Marina Sbisa (Oxford: Oxford University Press, 1979), p. 182.

¹¹ See, for example, Ludwig Wittgenstein, *Philosophical Investigations*, trans. G.E.M. Anscombe (Oxford: Basil Blackwell, 1953), §§ 23-24, 79-88.

universal principles that must be in place as soon as there is anything like describing what is real. Rather, “description” can mean many different things in different cases, and the grammar of description cannot be decided beforehand: one has to look and see what description amounts to in given cases.¹² However, this point does not by itself entail that grammar can be seen as an exclusively linguistic matter. Thus, when Wittgenstein notes that “*Essence* is expressed by grammar,”¹³ it is not at all clear that this is meant as a reductive claim, downgrading the notion of essence to a merely linguistic status. In fact, I do not think later Wittgenstein is any more inclined than his early self to situate grammar and logic on one of the sides of the language/reality dichotomy. Rather, he wants to say that this very dichotomy stands in the way of a proper understanding of the character and significance of grammar—and, of category distinctions.

Thus, there is a sense in which later Wittgenstein, like his early self, is closer to the Kantian tradition of critical metaphysics than to a Carnapian form of conventionalism. However, even if it is in some sense right to say that Wittgenstein, like a critical metaphysician, would not want to separate logic from reality, his most important achievement can become visible only if one qualifies this claim, by adding that the relevant notion of inseparability is crucially transformed in his later works. Indeed, according to later Wittgenstein, the very idea that we can say anything illuminating at all at a general level about what that inseparability-amounts to is itself a sort of philosophical misunderstanding. “Logic” and “reality” are not monoliths the “inseparability” of which can be stated and investigated in general terms; the ways they are interconnected will vary and will be clearly seen only if we look *in detail* at how language is used in concrete circumstances in human life. At such a level of detailed investigation, vaguely mysterious general claims such as “logic and reality cannot be separated,” “essence is expressed by grammar,” and “category-propositions tell us about the nature of things” will be replaced by much more specific reminders of how uses of language are intertwined with real-world surroundings in non-mysterious ways.

My main aim in what remains of this paper is to shed some further light on what this sort of point amounts to and why it is important. For this purpose, I shall make some use of a

¹² Ibid., §24.

¹³ Ibid., §371.

fine paper by Bernard Harrison from 1965.¹⁴ But first, I want to look at a more recent attempt to deal with the nature of categories and category mistakes.

3.

The features of category distinctions that I have discussed so far are all relatively well known and have been central to many debates over the last century. However, in recent discussions, there has been a peculiar tendency to ignore or downplay the traditional logical and metaphysical roles of category distinctions. The trend is instead to treat them from a sort of viewpoint deeply influenced by linguistics, approaching them as empirical phenomena in need of an explanation given at some level in the standard syntax-semantics-pragmatics taxonomy. Such an approach tends to misconstrue the question of whether category distinctions are only distinctions in language or whether they are also distinctions in reality by conceiving them as in effect *empirical* distinctions.

A striking example of this is Ofra Magidor's recent book, *Category Mistakes*, from 2013.¹⁵ Magidor's discussion proceeds from the idea that a category mistake is primarily given as a perceived "infelicity" that can be securely identified as a category mistake even before we have a satisfactory theoretical account of what it is that is infelicitous about it (CM 1). According to Magidor, "one can come to understand what is meant by 'category mistake' without an explicit definition of the term: its meaning can be learnt by ostension, via an ample range of relevant examples" (ibid.). As we shall now see, her approach leads to considerable obscurity when it comes to dealing with the question of whether category distinctions are only distinctions in language, or whether they are also distinctions in reality.

In chapter 2 of her book, Magidor attacks what she calls the "syntactic approach" to category mistakes, according to which "category mistakes are infelicitous because they are syntactically ill-formed or ungrammatical" (CM 25). Magidor discusses various arguments against such a view. I shall focus on how she thinks the syntactical approach runs into trouble due to its failure to account for how category mistakes depend on the interaction between language and extra-linguistic facts. Let us look at some detail at what Magidor has to say:

¹⁴ Bernard Harrison, "Category Mistakes and Rules of Language," *Mind* 74:295 (1965), pp. 309–25; henceforth CMRL, followed by page number.

¹⁵ Ofra Magidor, *Category Mistakes* (Oxford: Oxford University Press, 2013); henceforth CM, followed by page number.

Whether or not a sentence is classified as a category mistake depends not only on purely linguistic knowledge, but also on knowledge (or at least beliefs) of extra-linguistic empirical facts. Consider for example the following category mistake:

(19) This rock is thinking about the theory of relativity.

It is doubtful that our knowledge that rocks are inanimate, non-sentient, or incapable of thought is really part of our linguistic knowledge concerning the word “rock.” Rather, it seems that we have acquired the knowledge that, for example, monkeys can think but rocks cannot think by empirical investigation (possibly, empirical investigation that was carried out after one already had the relevant words in one’s lexicon.) But if so, then the infelicity of (19) cannot arise due to . . . purely linguistic, and in particular syntactic, features of the word “rock.” (CM 40–1)

I think the adequate philosophical response to this passage is to say that what Magidor is arguing is, in effect, that the sentence “This rock is thinking about the theory of relativity” is not a category mistake at all. For if she is right that we have acquired the knowledge that rocks cannot think by empirical investigation, then all philosophical motivation for calling the sentence a category mistake vanishes. The sentence might have the feel of a category mistake, all right; but if it turns out that it only expresses an empirical falsehood, then there is no good philosophical reason to classify it as such a mistake.

However, this is clearly not Magidor’s view of the matter. She does not take herself to be rejecting the idea that the sentence in question is a category mistake. Rather, she takes herself to be criticizing a certain explanatory model of how a category mistakes arise. To see how she understands the situation, consider how she continues her argument:

The point is perhaps strengthened if we consider sentences that in certain contexts would be deemed to be category mistakes, but which given additional background factual information may receive a different verdict. Consider for example:

(20) This priest is pregnant.

(21) This woman fathered my children.

(22) This machine is thinking about the theory of relativity.

One familiar only with the Catholic Church may well deem (20) to be a category mistake, but not so once one learns that some Christian denominations allow women to be ordained. (21) may sound highly infelicitous, until one realizes that this woman may have previously been a man who fathered my children, and then went on to have a sex-change operation. And (22) would probably be judged to be a category mistake by one who has not learnt of computers with sophisticated artificial intelligence capacities. (CM 41)

Again, the response to this by someone who is sensitive to the traditional philosophical interest in category distinctions, will be that what Magidor is in effect arguing is just that

(20), (21), and (22) are not category mistakes. From a philosophical viewpoint, to agree with her about the significance of the envisaged additional background information is precisely to acknowledge that those sentences were never category mistakes to begin with.

Now, interestingly enough, Magidor goes on by addressing what may appear as such a response. Her comments are instructive:

One might argue that (20)–(22) are not really category mistakes: rather, some erroneously judge them to be category mistakes because they lack the relevant empirical information. This proposal has some serious disadvantages (for one thing, it divorces the notion of a category mistake from the phenomenological quality that was used to characterize the phenomenon in the first place). But even if the proposal were adopted, it would entail that in order to discover whether or not a sentence is a category mistake, one must appeal to many extra-linguistic empirical facts. (CM 41)

Two things are striking about this passage. To begin with, Magidor repeats the idea that it is the empirical feel of infelicity, the “phenomenological quality,” that delineates the set of phenomena for which our theory of category mistakes should provide a unified account. This is the approach of an empirical linguist, and it deserves to be said again how very different it is from the viewpoint from which category distinctions and category mistakes can be seen as having genuinely deep philosophical significance.

Second, it is interesting to note Magidor’s taking it for granted that if I deny that (20), (21), and (22) are category mistakes, then I must take someone who judges them to be category mistakes to lack relevant empirical information. But again, from the sort of viewpoint that has traditionally motivated philosophers to make category distinctions, being mistaken about such distinctions is precisely *not* a matter of lacking empirical information. Consider Magidor’s example of the woman who fathered my children. Someone who thinks that this notion involves a category mistake, in the philosophically relevant sense of the term, has overlooked not merely the actual existence of sex-change operations, but their very conceivability.¹⁶ It may of course be that he is not imaginative enough to fathom the possibility of sex-change operations before they become reality, but this does not mean that his mistake is a merely empirical one. Similarly, if someone claims that it is a category mistake to say that a priest is pregnant, and I can oppose his verdict by referring to Christian

¹⁶ I follow Magidor here, who seems to have a biological rather than gendered conception of ‘father’ in mind. If we consider the concept of ‘father’ more generally, it would seem that gender transformations without sex-change operations suffice to allow for fatherhood. To avoid similar ambiguity I use ‘female’ rather than ‘woman’ below to indicate that the apparent category mistake regards conceptions of sex rather than gender..

denominations in which females can be ordained, then it would be sufficient to invoke the mere possibility of such denominations. The actual, empirical existence of female priests is not needed to reject his mistake; it suffices is female priests can be conceived at all.¹⁷

4.

Now, having criticized Magidor's approach, I should immediately add the following caveat. There *are* interesting things to say about how category distinctions are related to facts—empirical facts, if you like. And it is true that these relations may be hidden from view if one lays too much emphasis on the notion that category mistakes are mistakes that have to do with what is possible and conceivable rather than with what is actual, what happens to be the case. My main point in section 3 was that Magidor's treatment of category mistakes as empirical phenomena is what leads her to misconstrue how category distinctions are related to empirical facts. However, this is not to deny that there are other ways in which empirical facts can matter and indeed be crucial to how category distinctions work. The question of how empirical facts can matter in such ways is central to Wittgenstein's later philosophy.

To put Wittgenstein's later approach in a proper light, let me first return to the three different views of category distinctions that I identified earlier in this talk: (1) the anti-metaphysical, purely linguistic view, (2) the dogmatic metaphysical view, and (3) the critical metaphysical view. According to the anti-metaphysical view, category distinctions are only conventional distinctions within language and not in extra-linguistic reality. So, whereas it is perfectly appropriate to say things about how linguistic expressions are and are not to be used, the so-to-speak worldly counterparts to such linguistic statements are misguided. Thus, we might well say:

(L₁) In “. . . is taller than 1,500 feet,” “the Empire State Building” may well be inserted, but not “the number 2.”

¹⁷ In fact, this particular example is rather contrived. To begin with, I do not know of any important philosopher who would be tempted to classify the very notion of female priests as a category mistake. Moreover, someone who would seriously maintain that it *is* a category mistake would not thereby be denying that there might be denominations in which females function as so-called ‘priests’. Rather, he would say that such denominations are working with a categorically different [and perhaps theologically confused] conception of what the term ‘priest’ means.

Or, more generally, and assuming that one has made a list of ‘physical object words’ and ‘number words’ in the relevant language, one can say:

(L₂) In expression of the form “. . . is *n* feet tall” a physical object word may well be inserted, but not a number word.

By contrast, worldly or ‘material mode’ analogues are deemed inappropriate:

(W₁) The Empire State Building has a height, but the number 2 does not.

(W₂) Physical objects are spatial, whereas numbers are not.

Now, the problem with this view is that it cannot account for the kind of absurdity involved in category mistakes. We can of course imagine alternative conventions or languages such that joining the expressions “. . . is taller than 1,500 feet” and “the Empire State Building” are allowed. But that would *have* to mean that at least one of these expressions means something quite different from what it means in present English usage. And how can this “have to,” this “must,” be accounted for merely in terms of conventions of the envisaged sort? After all, mere conventions for how to combine orthographic units cannot give rise to a “have to” of the relevant kind, nor to the sort of genuine absurdities characteristic of category mistakes. I may have learnt certain rules for how to generate orthographic patterns for wallpaper decoration, say, and these rules may forbid the production of certain patterns. If those rules are broken, I may perceive the resulting deviant patterns as infelicitous. But they will not be absurd or nonsensical in the sense in which category mistakes are absurd or nonsensical. Deviant orthographic patterns are just unusual, and perhaps ugly or in other ways repellent or inconvenient.

One response here might be to thicken the notion of linguistic convention, so that it is taken to involve word-world correlations: one might think of the relevant conventions as *semantic*, as Ryle suggests (C 206). But then, it becomes really hard to explain why the worldly, material mode analogues would be in any way inappropriate when it comes to accounting for the necessities and the absurdities involved in categorial considerations. In fact, the idea that semantic conventions are what matters here leads quite straightforwardly to the conclusion that the fundamental explanation of the absurdity of category mistakes resides in the things themselves. For if we try to explain why “The number 2 is taller than 1,500 feet” is absurd by saying that “the number 2” refers to *the number 2*, and that “. . . is taller than

1,500 feet” means *being taller than 1,500 feet*, then that explanation will work only if the number 2 is indeed a sort of thing which has no height at all, and of which it would therefore be absurd to say that it is taller than 1,500 feet—or shorter, for that matter. So, it seems we then have to invoke precisely the sort of worldly, metaphysical considerations that the anti-metaphysical conception was meant to abolish.

At this point, one might recoil from the view of category distinctions as purely linguistic, and instead adopt its antithesis: the conception of dogmatic metaphysics. That is, one might argue that what is fundamental when it comes to category distinctions is indeed the metaphysical character of things themselves, and a language can do its job adequately only to the extent that its own structure is tailored to fit the independently present metaphysical structure of the world. According to this sort of view, when, as philosophers, we identify category mistakes, what we do is take a look at language use from sideways on, and see that it does not quite fit the world, so to speak: our employment of language is faulty in the sense that it allows combinatorial possibilities that are not allowed in the world itself. The metaphysical structure of the world is seen as an independently accessible touchstone by reference to which the meaningfulness of our thoughts and linguistic utterances can be assessed.

I will not here go through the many problems with this sort of view. Rather, what interests me is the possibility of a third alternative—what I have called a critical metaphysical conception. At first, it may be difficult to understand how there can be any room at all for such a third possibility: either category distinctions are conventions of language, and reality in itself is innocently uncategorized, or it is reality itself that contains these distinctions, and our linguistic conventions should be adjusted accordingly if we want to avoid nonsensical pseudo-propositions. Aren't these the two only alternatives? The critical metaphysician says no. According to her, there is no priority in either direction: the question of whether language or reality is the fundamental carrier of category distinctions is simply a bad question. Rather, the logical order of thought or language and reality is *the same*, and the very attempt to identify either the world or our means of representation without already presupposing such an order is misconceived from the start. And the order is really a *logical* order—not a merely conventional system for the manipulation of scribbles, ink-marks. Certainly, we can look at language use as the mere manipulation of orthographic units, but that is an artificial viewpoint, parasitic on our acquaintance with language from within its meaningful employment. From such an artificial viewpoint, language *as language*—as a means of describing what is in the world—will disappear out of sight. And once this artificial

viewpoint is mistakenly conceived as the fundamental one, it is of no help to try to retain meaning by adding semantic conventions of word-world correlation. As long as the words themselves are still conceived as mere ink-marks, such attempts to add meaning to them will only lead into the quagmire of dogmatic metaphysics. Rather, we must conceive of the logical order of language as an order that can be identified only if we already assume that the linguistic symbols are used to describe the world; the symbols can be identified as the symbols they are only if their descriptive function is taken for granted.

One way of putting this is to say that the anti-metaphysical conventionalist and the critical metaphysician have radically different conceptions of what sort of *syntax* gets described in statements such as (L₁) and (L₂). The anti-metaphysical conventionalist thinks of that syntax along the lines of Carnap, Tarski,¹⁸ and others—basically, as rules for the manipulation of orthographic units that are in themselves meaningless. The critical metaphysician, on the other hand, conceives of the syntax in question as a syntax the rules of which presuppose that the units whose combinatorial possibilities get presented are already meaningfully related to reality.

So, the critical metaphysician will agree with the anti-metaphysical conventionalist that there is no way of exploring category distinctions from sideways on—the investigation of category distinctions will always be an investigation into the structure of our means of representing the world. However, she will disagree with the anti-metaphysical conventionalist about the character of statements such as (L₁) and (L₂). According to the anti-metaphysical conventionalist, such statements are statements about how we conventionally combine words *qua* mere orthographic units. For the critical metaphysician, by contrast, statements such as (L₁) and (L₂) are logico-conceptual reminders that presuppose that the units whose use gets described are units already in meaningful employment. And this means that a wedge cannot be driven between the structure related by such statements and the structure of the world that we represent in language. Again, the logical structure of language and the metaphysical structure of reality are really *the same* structure—and thus there just is no question of to what extent the one “corresponds” to the other.

¹⁸ See, for example, Alfred Tarski, “The Concept of Truth in Formalized Languages,” in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, trans. J.H. Woodger, ed. John Corcoran (Indianapolis: Hackett, 1983), pp. 152–278.

5.

Now, an interesting question is whether the author of the *Tractatus* is a critical metaphysician in the sense just described. One may sense that the early Wittgenstein does not quite fit my scheme: after all, as I said earlier, he thinks a logically adequate notation will not allow statements such as (W₁) and (W₂) at all, and that conceptual distinctions will show up in such a notation as rules for the use of signs. But in fact, I am inclined to think that early Wittgenstein took himself to have thought through a critical metaphysical conception to the end, and to have arrived at its most adequate formulation. His key idea here, I believe, is that if we consider the nature of the relevant syntax and of the units whose use constitutes that syntax, we will have to conclude that there is no room here for notions such as may and may not, can and cannot, and must and must not. Rather, there just *is no such thing* as, say, inserting the expression “the number 2” in the empty slot of “. . . is taller than 1,500 feet.” Of course, we can place the orthographic unit “the number 2” right before the orthographic unit “is taller than 1500 feet,”[‡] and we may even give that combination a sense, if we like. But there just is no such thing as combining the two expressions while at the same time retaining their usual meaning—the one they normally have in, say, “The number two is prime” and “The Empire State Building is taller than 1500 feet.” Consequently, the very idea that there is some sort of normativity or necessity to explain here by reference to reality “itself”[‡]—the very idea that the world somehow favors or necessitates one way of using language as opposed to another—evaporates,[‡] and reveals itself to be just one more variety of dogmatic metaphysics.

I will not say more about early Wittgenstein here. Instead, I want to spend what remains of this paper gesturing at how I think Wittgenstein’s later thought can be related to all this. I believe later Wittgenstein is no closer to either dogmatic metaphysics or to anti-metaphysical conventionalism than his early self. I suggested above that even later Wittgenstein inherits something from critical metaphysics. And I think that is true: he retains the idea that philosophically relevant investigations of language are investigations from within that meaningful employment with which we are already acquainted as competent language users. Such investigations remind us of what we already know. These investigations do not explain anything new—for example, what needs to be added to meaningless scribbles and sounds to achieve meaningful words.

However, what I think is most interesting and radical about later Wittgenstein’s approach to the sort of issues I have discussed in this essay is that he refuses to deal with the problems at the level that characterizes all three conceptions so far described—anti-

metaphysical conventionalism, dogmatic metaphysics, and critical metaphysics. As I have described these three positions, they all pose and try to answer the *general* question of how we should conceive of category distinctions: as only or fundamentally linguistic, as primarily in the world, or as distinctions in a structure shared by language and world? Later Wittgenstein, by contrast, thinks that this sort of general approach is superficial and unhelpful. With regard to my sentences (L₁), (L₂), (W₁), and (W₂), I think he would say that *none* of these sentences helps us really understand what we need to understand when we puzzle over category distinctions and category mistakes. They all remain at the surface by making it seem as if the relevant logic, the relevant grammar, can be reduced *either* to a matter of how linguistic symbols may be combined, *or* to a matter of what kind of properties things themselves can and cannot have. But neither of these sorts of statements take us anywhere. Rather, they merely reiterate what really puzzles us: the mays and may nots, the cans and cannots, or even the Tractarian, there is no such thing as combining. . . . What we need to understand is to what these apparent possibilities and restrictions and “there being no such thing as” actually amount—that is the puzzle of category distinctions and category mistakes. And in order to really understand that, what we need to do is go beyond that surface and consider what particular category distinctions amount to in real-life linguistic practice. We need to look in detail at the use of the relevant words,⁵ and then the distinctions and their significance will no longer look so mysterious.

In his illuminating but all-too-neglected paper, “Category Mistakes and Rules of Language,” Bernard Harrison approaches the issue of category distinctions and category mistakes in this sort of Wittgensteinian fashion. Discussing Ryle’s example of the child who is puzzled by the fact that the equator can be crossed but not seen, Harrison asks us to consider a couple of statements parallel to my (L₁) and (L₂), and (W₁) and (W₂), respectively:

- (L₃) “Equator” cannot be inserted in “He saw the . . .” and related sentence-frames.
 (W₃) The equator cannot be seen.

With regard to these statements, Harrison writes:

Each of these statements asserts the existence of a puzzling situation, and each of them, by its style of formulation, makes a suggestion about where the locus of the puzzlement may ultimately be found . . . ; but none of them tells the child what it really wants to know, namely, *why* the equator cannot be seen, or to use Ryle’s terminology, *why* “equator” cannot be inserted in “he saw the . . . looming over the horizon” and related sentence-frames. In the case of the equator, it is perfectly easy to

give the required explanation by showing the child how cartographers have come to use the word “equator.” But such an explanation will consist primarily of propositions about map-making, and it is difficult to see why these propositions should be called category-propositions. (CMRL 311–2)

And he goes on by saying:

We can give explanations, like the cartographer’s explanation of why the equator can be crossed but not seen, which seems to go beyond logic in the narrow sense. The cartographer’s explanation does not simply describe the behavior of a word in various verbal contexts: it gives what one feels tempted to call an explanation of the function in human social life of the word in question The cartographer’s explanation shows us, as it were, in what sense the logic of ‘equator’ stems from the nature of things and in what sense from the nature of our linguistic conventions: for this reason I am inclined to think that a study of what is involved in such explanations is a necessary step toward answering the questions “What sort of linguistic rules do category-mistaken sentences break?”, “Are category propositions ‘About language’ or ‘About the Real’?” (CMRL 312)

Here we can begin to see how category distinctions *are* related to facts—empirical facts if you like, general facts of nature that could have been otherwise. The cartographer’s use of the word “equator” would not be what it is if the earth was not a sphere rotating around a fairly stable axis; if the earth had instead been quite irregularly shaped, and had turned and twisted haphazardly in space, nothing like our term “equator” would have any useful function. Harrison spends a lot of time on another example: “Thursday is sleeping” (CMRL 316). By first reminding us of some characteristic features of our use of the names of weekdays, and of how that use depends on seemingly trivial facts such as the fact that periods of light are regularly followed by periods of darkness, and then, in a similar vein, reminding us of what is involved in saying of persons that they are asleep or awake, Harrison is able to clarify, in an utterly non-mysterious manner, why there is no such thing as Thursday’s being asleep, or Thursday’s being awake (CMRL 318).

Similarly, one could describe in as much detail as would be required our practices of talking about and measuring how tall something is—a description that would have to mention basic facts about measuring-rods and similar instruments, and remind us of how comparing physical lengths presupposes that the things one compares have a suitably rigid physical shape, and so forth, and then one could compare this with our practices with number-words, and thereby show what it means to say that it does not make sense to ask how tall a number is.

There is just one thing in the second quotation from Harrison that I am uncomfortable with, and that is how it ends, when he says about such concrete, detailed descriptions that

they are necessary steps toward answering the questions, “What sort of linguistic rules do category-mistakes break?” and “Are category-distinctions ‘about language’ or ‘about the real’?” Here, I think Harrison chickens out: for now he suddenly starts treating the detailed descriptions of language use as mere preparations for answering the great philosophical questions about category distinctions and category mistakes (CMRL 312). Whereas I take the truly radical Wittgensteinian insight to be that giving these detailed descriptions is sufficient—those descriptions make it superfluous to go back to the grand question, “Are category propositions about language or about the real?” That very dichotomy is false. What the detailed descriptions show is that category propositions are about both language and reality, and language and reality intermingle in various ways within our life with words. There is no one great answer to this grand question, but a lot of detailed and non-mysterious answers that will vary depending on what particular category distinction is at issue.

So, later Wittgenstein is not an anti-metaphysical conventionalist, a dogmatic metaphysician, or a critical metaphysician, as I have described these different conceptions in this paper. Rather, he helps us see that the proponents of these three positions are all held captive by one grand scheme that stops them from looking at the sort of details they would need in order to really get rid of what puzzles them about categories and category mistakes. Not that Wittgenstein would deny that numbers lack height, that the equator cannot be seen, or that Thursday can neither be awake nor asleep. These claims are all perfectly correct, as far as they go. The problem is that they do not go very far, and that they may stand in the way of genuine philosophical clarity.