

## THE DISCOVERY THAT PHOSPHORUS IS HESPERUS: A FOLLOW-UP TO KRIPKE ON THE NECESSITY OF IDENTITY

M.J. GARCÍA-ENCINAS

encinas@ugr.es

Universidad de Granada

**ABSTRACT.** It was an empirical discovery that Phosphorus is Hesperus. According to Kripke, this was also the discovery of a necessary fact. Now, given Kripke's theory of direct reference one could wonder what kind of discovery this is. For we already knew Phosphorus/Hesperus, and we also knew that any entity is, necessarily, identical to itself. So what is it that was discovered? I want to show that there is more to this widely known case than what usual readings, and critics, reveal; and I want to show this under the Kripkean pattern that what was discovered is the plain fact of identity that Phosphorus is, and necessarily is, Hesperus. Moreover, I will show how both Kripkean theses, the necessity of identity and the direct theory of reference, should not be kept apart to understand the kind of discovery this is. In particular, understanding the kind of discovery we made will help us see how intuition is displayed in metaphysics, and how metaphysical impossibilities, and metaphysical modality in general, can be discerned by reason and separated from other modalities. The study of this discovery will help us see, in a line, how identity belongs to the inner and most profound structure of reality and to the most profound structure of cognition and language.

**Keywords:** identity; discovery; fact; necessity; intuition; metaphysics

How to cite: García-Encinas, M.J. (2017), "The Discovery that Phosphorus Is Hesperus: A Follow-up to Kripke on the Necessity of Identity," *Analysis and Metaphysics* 16: 52–69.

*Received 7 October 2016 • Received in revised form 14 February 2017*

*Accepted 14 February 2017 • Available online 27 February 2017*

### I

*The discovery case.* Kripke has argued for these well-known ideas:

(1) *The direct reference theory for proper names.* Proper names directly refer to their bearers; they do not stand for concepts; they do not have any sort of Fregean senses, nor abbreviate Russellian descriptions. "Hesperus" does not mean the first star you see in the evening, nor the planet we call Venus; it does not stand for a

concept, or for any of your favorite descriptions of Venus; it does not mean anything. The sole and full semantic role of “Hesperus” in a statement is to refer, directly, to Hesperus.<sup>1</sup>

(2) *The necessity of identity.* Identity, a metaphysical relation, holds of necessity. So Hesperus is necessarily identical to Phosphorus, if they are identical.

(3) *Identity facts can be known a posteriori.* It was an empirical discovery that Hesperus is Phosphorus, and Venus. Pierre could discover a posteriori that *Londres* is London. Quine learned that Tulio is Cicero. Schrödinger surprised Quine when claiming that Mount Everest was Mount Gaurisankar, etc. These facts of identity were established a posteriori.

But when put together, (1) and (3) seem to raise an unsettling question: *what* is it that one discovers or learns in cases like these? If proper names do not have senses, nor hide any kind of description, if they just name a given particular entity, what do we learn when we learn that Hesperus is Phosphorus, that *Londres* is London, that Tulio is Cicero? What did Schrödinger claim that surprised Quine? Thesis (2) seems to add to the mystery. For if identity is necessary, what is discovered includes this necessity. And if these statements are necessary, their proper names referring to the same entity, is not their necessity a sign of a redundancy that clashes with (3)? How could one have been wrong or ignorant of them?

I will consider four main answers to the discovery question; I will reject the first three, and defend the fourth – the only one that wants to be faithful to the three Kripkean theses above. I call it the metaphysical solution, as it highlights the idea that the discovery is the discovery of a metaphysically necessary matter of *fact*, and I confront it to other answers to the question, where linguistic, conceptual, or logical aspects play the main part. This occurs in part II of the paper. But in the process of understanding the kind of discovery we made, some crucial ideas concerning the epistemology of modality will also be disclosed, and the role that (2) is playing in the discovery will be highlighted. For the discovery of a fact of identity is a constant and ordinary kind of empirical discovery. It is also a direct immediate discovery, with no other intermediary standing between our apprehension of the fact and the entity in question. Identity, a relation necessarily hooking one entity to itself, displays itself in every empirical discovery of this kind. So its apprehension is an act of intuition where identity is seen to fit within the basic and necessary structure of reality in a particular fact. Part III is devoted to these ideas.

## II

*The metalinguistic solution.* One possible answer to the discovery question says that when we empirically learn, say, that Hesperus is Phosphorus, what we learn is that we had tagged the same object twice. This was Quine’s position. Kripke quotes Quine against Marcus’s conclusion that identity statements between proper names are a priori, so that in a perfect dictionary their entries would mutually refer

to each other.<sup>2</sup> Quine replies to Marcus that identity statements can be known a posteriori, in this sense: “We may tag the planet Venus some fine evening with the proper name ‘Hesperus.’ We may tag the same planet again someday before sunrise with the proper name ‘Phosphorus.’ When, at last, we discover that we have tagged the same planet twice, our discovery is empirical, and not because the proper names were descriptions” (1971/2011: 6).

Quine, as read by Kripke, claims that when we discovered that Hesperus is Phosphorus, what we empirically discovered is a fact about our use of language, i.e., that we had two different names for the same entity. Kripke immediately argues against this: we did make (against Marcus) an empirical discovery, but (against Quine) this is not one of a linguistic kind. We did not discover a fact about the name “Hesperus” or about the name “Phosphorus;” rather we discovered a fact about Hesperus/Phosphorus itself. Moreover, Hesperus/Phosphorus could have held the same name, and we could have made the same empirical discovery concerning their identity.

It is worth noting that Kripke also rejects a metalinguistic position when wondering about statements of existence. To claim that Moses did exist, and unicorns and Holmes do not, is not to say that “Moses” has a referent, but “unicorn” or “Sherlock” do not refer. Kripke is quite clear that this proposal fails the counterfactual test: “Neither in the case of the name ‘Moses’ nor in the case of ‘Sherlock Holmes’ does this metalinguistic translation give an analysis that would apply to counterfactual situations also. Counterfactually speaking, Moses might have existed even though the name ‘Moses’ had no referent. This would be the case if neither he nor anyone else had ever been called ‘Moses.’ It is also true that the name ‘Moses’ might have had a referent, where that referent might not have been Moses” (1973/2011: 70).

Existential statements are existential claims about entities, not linguistic claims about names.<sup>3</sup> To say that Moses existed is to say that there was someone, Moses. This claim could be true even in the case Moses had not been named “Moses,” or in the case someone other, or nobody, had held the name “Moses.” As for negative claims, such as “Holmes does not exist,” they merely mean that there is no proposition expressed by “Holmes exists,” and this is as well all there is to the falsity of this last sentence. There is no truth of the matter about Holmes.<sup>4</sup> You could discover that Holmes does not exist after many years believing that he did, and what you would discover is not that ‘Holmes’ does not refer, but something about our world. Also, when a child learns that Santa does not exist he learns that there is no individual that is Santa, not that “Santa” does not refer. The child will use the name, not mention it, to express his acquired knowledge that there is no fact in the world that includes Santa.<sup>5</sup>

Whatever the merits or demerits of Kripke’s own proposals over existence and empty names, the problems raised in these fields seemed like a golden opportunity for the adoption of a metalinguistic theory, but Kripke always left it to one side. In a parallel way, metalinguistic new knowledge, even if it were also obtained or

implied in the discoveries of relations of identity, does not solve the discovery case. For identity statements are not statements about the linguistic role of their expressions. The discovery is not linguistic in any sense, nor it has to do with the role of names in any statement of identity. “Hesperus” does not abbreviate anything of the form of the entity named by “Hesperus.” This would go as far as betraying (1) itself, apart from being a circular analysis of the role of “Hesperus.” Moreover, Hesperus could have been discovered to be Phosphorus, even if it had had no name, or even if some other celestial body rather than Venus had held the name “Hesperus.” Names belong contingently to their bearers, so adopting this view on the discovery would also betray (2). Whatever it is that we discovered when we discovered that Hesperus is Phosphorus, we did not primarily discover that “Hesperus” and “Phosphorus” named the same entity.

*Descriptivist and Fregean solutions.* Once and again Kripke denies different versions of the Frege-Russell view about proper names. In the Fregean theory of language and logic, identity statements are declared contingent when they are cognitively significant, i.e., when they are not tautological but rather seem to provide significant or new information. Frege explains this significance by proposing different senses, or different Husserlian modes of presentation of the same entity, which appropriate different names contribute to the proposition. Thus, according to Frege, when we hear that “Hesperus is Phosphorus” we learn or understand that the evening star is the morning star, or something of the like. If you rather prefer a Russellian theory for the semantics of proper names, these will be disguised forms of definite descriptions. Following Russell, you could say that what Schrödinger claimed, in the Everest case, is that the highest mountain seen from Tibet is the highest mountain seen from Nepal.<sup>6</sup> According to both, Frege or Russell, proper names do not directly refer to their bearers, but their referential role is mediated by concepts or descriptions, for which they “really” stand in the sentence.

Now, even though Kripke introduces his theory of reference in direct confrontation to views such as Russell’s or Frege’s, it is still possible to think of forms of descriptivism or conceptualism as a useful complement to the Kripkean program, that perhaps in some modified version could help in explaining the cognitive value of certain identity statements attending to the information *previous* to the discovery of their truth or falsehood. Kripke himself seems to have left some clues in this direction in *Naming and Necessity*. Consider the case of heat. There certainly was a time before the scientific discovery that heat is molecular motion. At that time, we people did not know that heat was molecular motion, but we knew that hot things produced certain sensation in us. We seem to have been in an epistemological situation where heat was thought of, or identified as, the phenomenon that produces certain sensation in us. Of course this was a contingent property of heat, given that heat could have not produced that sensation. But the relevant fact is that, before knowing that heat is molecular motion, we were in a situation where heat was *epistemologically* identified with the phenomenon that produces certain sensation

and, according to Kripke, this fact explains the apparent contingency of our discovery when we discovered the necessary fact that heat is molecular motion.

Thus, interpretation is open: When we discovered that heat is necessarily molecular motion, what we discovered is that the entity that contingently produces certain sensation in us is (necessarily) molecular motion. (Same as when we *conceive* that heat is not molecular motion: what we conceive is that what produces certain sensation in us is not molecular motion).<sup>7</sup> Equally, when we discovered that Hesperus is Phosphorus, what we discovered is that the celestial body that is, contingently, seen in the evening is (necessarily) Phosphorus.

But I think this interpretation should be resisted. If it were right, the senses, descriptions, or properties, if there were any, under which we could think of Hesperus/Phosphorus before the discovery, or the descriptions which could have fixed the reference of “Hesperus”/“Phosphorus,” would be part of the discovery; but they are *not*. No sense, description, or property of Hesperus/Phosphorus belongs in the a posteriori proposition that Hesperus is Phosphorus. And Kripke insists that this reading would affect his semantics for proper names. *Moreover*, he writes: “...given the evidence that someone has antecedent to his empirical investigation, he can be placed in a sense in exactly the same situation, that is a qualitatively identical epistemic situation, and call two heavenly bodies ‘Hesperus’ and ‘Phosphorus,’ without being identical” (1972/80: 104).

We could have had exactly the same information about Hesperus/Phosphorus in a possible situation in which it was discovered that there were two planets, or two distinct stars, one seen in the morning and the other before sunset, one called “Phosphorus” and the other “Hesperus.” Sure, this would not be a situation in which Hesperus (our Hesperus – what else?) is not Phosphorus. But if this situation and the actual situation where we discovered that Hesperus is necessarily Phosphorus are epistemologically identical before the correspondent discoveries took place, then, when it was discovered that Hesperus is Phosphorus, and “when” it could have been discovered that some other entity called “Hesperus” is not Phosphorus, the properties in question (i.e., the properties, senses, descriptions, etc., if any, by which we had imagined Hesperus/Phosphorus, or the other body) are irrelevant or uninformative about *what* was discovered.

As another example, consider Kripke’s particular puzzle about belief. One of the main points in the well-known paper is that appealing to senses, descriptions, etc. does not solve the puzzle. Pierre can have exactly the same beliefs about *Londres* and about London (except for that *Londres* is pretty, and London is not pretty), but he does not know that *Londres* is London. Pierre could even have fixed the reference of “*Londres*” and the reference of “London” by exactly the same properties, unknowing that *Londres* is London. So when he discovers, if he ever does, that “*Londres* is London” the idea that what he discovers is that *the so and so* is *the so and so* is worthless:

The puzzle can still arise even if Pierre associates to ‘*Londres*’ and to ‘London’ *exactly* the same *uniquely identifying* properties. How can this

be? Well, suppose that Pierre believes that London is the largest city in (and capital of) England, that it contains Buckingham Palace, the residence of the Queen of England, and he believes (correctly) that these properties, conjointly, uniquely identify the city. (...) Earlier, when he spoke nothing but French, however, he associated *exactly* the same uniquely identifying properties with ‘*Londres*.’ He believed that ‘*Londres*,’ as he called it, could be uniquely identified as the capital of England, that it contained Buckingham Palace, that the Queen of England lived there, etc. Of course he expressed these beliefs, like most monolingual Frenchmen, in French... (1979/2011: 148; original italics.).

Pierre could have fixed the reference of “London” and of “*Londres*” by the same properties, he could associate exactly the same properties to “London” and “*Londres*,” and yet he could remain unaware of the fact that London is *Londres*. In such case, adding the claim that Pierre would discover that the so and so is the so and so if he ever discovered that *Londres* is London, is not adding to his discovery. And then, senses or descriptions of *Londres* or London need not be part of his finding that London is *Londres* when these senses or descriptions differ.<sup>8</sup>

In the situation of ignorance of the identity between Phosphorus and Hesperus we could have had the very same senses/descriptions/thoughts about Hesperus and about Phosphorus, and still believe that they were two or be unaware that they were one. No version of descriptivism can explain *what* we would have then discovered in discovering that Phosphorus and Hesperus were one and the same entity.

Not only our favorite descriptions or senses are not part of our discovery, but it is also hard to see how these kinds of approaches could help in explaining that the discovery is the discovery of a necessary truth. Epistemic possibility, in terms of conceptual conceivability, is not a reliable guide to metaphysics modality. Whether it is in the form of images, concepts, propositions, ... we can conceive what is impossible. We can conceive that heat is not molecular motion; we can conceive that the sun is a god, that  $\pi$  is not 3.1416, and so on. And when conceiving the impossible, we do not conceive that *other* possible entities with the appropriate properties stand in other relations of identity. It is possible, to be sure, that most of the properties that heat has (such as that of producing certain sensation in us), being contingent, could have belonged to some other stuff that is not molecular motion. It is possible that most of the properties of the sun, being contingent, could have belonged to a miraculous entity. These are possible, and conceivable, situations; but these possibilities are not what we conceive when we conceive that heat is not molecular motion, or that our sun is a god.

Conceivability is just not the guide to modal knowledge. So conceptual conceivability cannot explain why discovery cases of identity are discoveries of necessary truths. The next section, it is my aim, shows the relevance of the claim that discoveries of identities are discoveries of necessary truths, and the role that this necessity plays in the philosophical uneasiness of their discovery.

*The logical solution.* Even if not directed to it, one answer to the discovery case could say that the discovery that Hesperus is Phosphorus is the discovery that any

fact, or any truth, about Hesperus must be a fact or a truth about Phosphorus, on pain of contradiction. Pierre is the usual suspect. If “London” and “*Londres*” refer to the same entity, when Pierre claims, at the same time, that London is not pretty and that *Londres* is pretty, he is, unknowingly, contradicting himself. As far as Pierre remains unaware of the identity between London and *Londres* he could be saved from irrationality; but he is, inadvertently, assenting to contradictory claims. (And from there on the discussion centers on how contradiction and rationality should be reconciled.)

Irrationality apart, the salient point for our own discussion is the claim that Pierre contradicts himself. If he only knew that “London” and “*Londres*” have the same semantic content, he would be aware of his holding contradictory beliefs, and he would somehow adjust them to consistency. As Pierre ignores that London and *Londres*, in his different beliefs, refer to the same entity, the problem with Pierre is usually known, since Boghossian (1994), as failure of epistemic transparency. Thus, reject epistemic transparency and the puzzle is solved. In other words, reject that, given that Pierre holds some beliefs about London/*Londres* and that he does understand the claims he makes about the city, Pierre should know a priori that “London” and “*Londres*” have the same semantic content, and you will understand why Pierre, inadvertently, contradicts himself.<sup>9</sup>

Clearly, Pierre does not know that London is *Londres*. So it seems obvious that he cannot know a priori that London is ugly and *Londres* is ugly have the same semantic content. But this epistemic opacity is not a reason for unrecognized contradiction. Note first that, even if not knowing that London and *Londres* have the same semantic content, Pierre *does* know what “London” and/or “*Londres*” stand for. To maintain that, in some of his claims, he does not refer to the city, for instance, because he has only seen parts of it, would imply that we are unable to refer to almost anything. There is no significant difference between Pierre’s references to London and the references to London of some other Londoner; and there is no significant difference between Pierre’s references to *Londres* and the references to *Londres* of some other French or French-speaker. So Pierre does properly refer to the same city by “*Londres*” and “London,” even if he does not know that London is *Londres*. But second, and more to the point, Pierre’s claims of belief about London and about *Londres* are *not* contradictory. Let’s see.

Pierre does never assent to anything like “London is pretty and London is not pretty.” In fact, we are pretty sure that he would deny such a belief. Given Pierre’s logical abilities, not only Pierre cannot infer the contradiction that London is and is not pretty, but he should be accused of logical fallacy if he were to infer this contradiction from his separate beliefs that London is not pretty and that *Londres* is pretty. In the same way, Kripke warns us against falsely accusing Pierre of logical fallacy if he were to infer from “London is not pretty” plus “If New York is pretty then *Londres* is pretty,” that New York is not pretty (Kripke, 1979/2011: 146).

So does Pierre learn, if he ever does, when he learns that London is *Londres*, that he had been contradicting himself until the moment of his discovery? I do not

think so. Imagine that I hear a cracking noise in the kitchen. I suddenly remember to have left my favorite cup of coffee at the edge of the sink, and I think that it has broken. Let's accept for the sake of the argument that cups are necessarily made of their material, and that I know this metaphysical principle. It is physically impossible that, under normal circumstances, like these, things made of methacrylate break when they fall from sinks, and I am also acquainted with this law. However, when I learn afterwards that my cup is made of methacrylate, and not broken, I do not learn that I was contradicting myself. Had I known that it was made of methacrylate, I should perhaps have deduced that it was not broken. But I do not discover, when I learn that it was made of methacrylate, that I had been contradicting myself when thinking that it was broken. Now, apart from the fact that the relevant theses involved here (i.e., that my cup is made of methacrylate, or that, in given circumstances, it is impossible that methacrylate things break when falling) are clearly metaphysical in character, this case is not much different from Pierre's.

Pierre knows the principle that same things have same properties  $((x) (y) (x = y) \rightarrow (Px \rightarrow Py))$ .<sup>10</sup> He believes that *Londres* is pretty ( $Pa$ ) and that London is not pretty ( $\neg Pb$ ). But he does not know that  $a = b$ . Given the principle, and the fact that London is *Londres*, it is impossible that *Londres* is pretty and London is not, but Pierre is unaware of this impossibility because he does not know that *Londres* is London. And no logic can give him the truth of this proposition of identity. So if he became aware of the impossibility that *Londres* is pretty and London is not, he would not become aware of a contradiction. It is plainly not logically necessary that *Londres* is London, that  $a = b$ , or that any entity is identical to itself – as it is not logically necessary that my cup is made of methacrylate.

We can imagine that Pierre also knows the additional thesis that identity is necessary. But again, the principle that same things are necessarily the same things  $((x) (y) (x = y) \rightarrow L(x = y))$  does not make Pierre's claims contradictory. This principle does not tell Pierre that London and *Londres* are its instances. So Pierre does not learn that he was contradicting himself when believing that London is not pretty but *Londres* is. So when discovering that London is necessarily *Londres*, he will not realize that it is contradictory to deny that London is *Londres*, for it is not. But if it is not contradictory to deny that London is *Londres*, it is not contradictory to claim that London is not pretty, but *Londres* is.

(Pierre would not realize either that *Londres* is necessarily identical to itself: he already knew that. At least we can suppose that everybody, including Pierre, agrees a priori that every entity is necessarily identical to itself.)

It is not contradictory to deny that London is *Londres*; and it is not contradictory to claim both, that London is not pretty but *Londres* is. It is not contradictory to deny that Hesperus is Phosphorus; and it is not contradictory to claim both that Hesperus is the evening star and Phosphorus is not the evening star. People did not hold an incredible quantity of unknowingly contradictory beliefs before discovering that Hesperus is Phosphorus. They did hold plenty of false beliefs (that Hesperus is not the morning star, that Phosphorus is not the evening



star; and many others of the kind), and they also held the impossible belief that Hesperus is not Phosphorus. But an impossible belief needs not be impossible because it is contradictory. Deny that Phosphorus is Hesperus, and you would be denying a necessary claim, but not a claim of logic. For the claim is necessary in so far as it expresses the necessary relation of identity between an object and itself. And this is primarily a question of metaphysics.<sup>11</sup>

*The metaphysical solution.* When we learn that Hesperus is Phosphorus we learn a, metaphysically necessary, matter of *fact*: we discover that what had been previously supposed to be two distinct entities are, in fact, one and (necessarily) the same one. It follows from this that we *also* discover that “Hesperus is Phosphorus” and “Hesperus is Hesperus” express the same proposition; we discover that Hesperus is Phosphorus and Hesperus is Hesperus is the same fact. And there are many discoveries, of different kinds, following the discovery of the identity between Phosphorus and Hesperus, such as that we had tagged the same object twice, that we were wrong in believing that Phosphorus was not the evening star, etc. But what we primarily discover is that Hesperus *is* Phosphorus. The discovery that Hesperus is Phosphorus, that *Londres* is London, or that Mount Everest is (not) Mount Gaurisankar, ... has cognitive significance because we did not know that Phosphorus and Hesperus (or any of the others) was the same entity. We thought that they were two entities, when “they” was only one.

Before getting into the philosophical relevance of this discovery, there is a question still concerning the previous section that I have not wholly answered: If “Hesperus is Phosphorus” and “Hesperus is Hesperus” have the same semantic content, how can it be not contradictory to deny “Hesperus is Phosphorus” but it is contradictory to deny “Hesperus is Hesperus”?

I should note, in the first place, that Kripke did not declare that “Hesperus is Hesperus” and “Hesperus is Phosphorus” express the same proposition, or have the same semantic content. In fact, he left the question unanswered. This is what he said:

Russell does seem to have held such a view for ‘logically proper names,’ and it seems congenial to a purely ‘Millian’ picture of naming, where only the referent of the name contributes to what is expressed. But I (and for all I know, even Mill) never intended to go so far. My view that the English sentence ‘Hesperus is Phosphorus’ could sometimes be used to raise an empirical issue while ‘Hesperus is Hesperus’ could not show that I do not treat the sentences as completely interchangeable. (...) How this relates to the question what ‘propositions’ are expressed by these sentences, whether these ‘propositions’ are objects of knowledge and belief, and in general, how to treat names in epistemic contexts, are vexing questions. I have no ‘official doctrine’ concerning them, and in fact I am unsure that the apparatus of ‘propositions’ does not break down in this area. [Kripke refers to Pierre’s puzzle in the note that follows this last sentence, and ends:] ... No firm doctrine regarding the point should be read into my words (1980: 20–21).

I want to pick up the gauntlet and claim that to follow this strong version of Millianism is the best way to go, and to go in continuity with a Kripkean program. Thus, I have claimed that “Hesperus is Phosphorus” and “Hesperus and Hesperus” express the same proposition when “Hesperus” and “Phosphorus” refer to the same entity. They are not, however, completely interchangeable. And this leads us back to the original question: If “Hesperus is Phosphorus” and “Hesperus is Hesperus” have the same semantic content, how can it be not contradictory to deny “Hesperus is Phosphorus” but it is contradictory to deny “Hesperus is Hesperus”? And, adding Kripke’s own concern, cannot “Hesperus is Hesperus” be used to raise an empirical issue?

I think that the answer to these questions is straightforward once you move (and this movement seems to me to be a very natural movement within the Kripkean program) from the realm of meaning to the realm of metaphysics, that is, if you think that for two sentences to have the same semantic content is for them to express the same proposition or *fact*. Thus, when you discover that “two” facts is the same one, you do not discover that two thoughts, or two Fregean senses, are thoughts about the same reality. Statements with truth-values do not express Fregean thoughts, but Russellian propositions or facts: true statements *state* that certain properties belong to certain entities, or that certain entities stand in certain relations, as in the case of identity statements.

Hence to deny that two particular facts are the same fact, when they really are the same one, is to state something that is metaphysically impossible, but it is not a logical (not a conceptual) contradiction. Remember that it is not logically impossible, even if it is metaphysically impossible, that Hesperus is not Phosphorus. The impossibility that Hesperus is not Phosphorus is not a question of formal logic, but of metaphysics. It is metaphysically impossible that an object is not itself; and Hesperus is an object. In *this* sense, “Hesperus is Hesperus” and “Hesperus is Phosphorus” express the same fact or proposition. It is impossible that Hesperus is not Hesperus/Phosphorus; but it is not contradictory to deny it. It is also impossible that the fact that Hesperus is Hesperus and the fact that Hesperus is Phosphorus is not the same fact; but it is not contradictory to deny it.<sup>12</sup> However, the impossibility that Hesperus is not Hesperus could also be considered a question of logic, if it seems impossible that a logical system denies that Hesperus is Hesperus. In this *other* sense, it is also contradictory that Stringer (an invented name) is not Stringer. In a sense that is wholly irrelevant to any metaphysics, it could be logically contradictory to deny that Stringer is Stringer – instead of “Stringer” or “Hesperus,” I could have used any symbol whatsoever. But the sense in which it would be logically contradictory to deny that Stringer is Stringer has nothing to do with Stringer: there is no fact about Stringer, there is no expressed proposition by “Stringer is Stringer,” for Stringer does not exist. (Remember the case of Holmes.)

So we can discover that, when we thought that we had empirically consigned two facts, “they” was really only one (and necessarily the same one). That Phosphorus is the evening star and that Hesperus is the evening star is the same fact, but it is

not logically contradictory to deny that “they” is the same fact. In the same sense, we can discover that, when we thought that we had empirically consigned one fact, they were really two (and necessarily distinct) facts. It is impossible that, if they were two, they “was” the same one, but it is not contradictory:

I say “Hello” every morning to Anna when I arrive at the Faculty. Every evening I say “Bye” to Anna when I see her jogging at the park. But one day I discover that Anna and Anna (in fact, she is called “Manuela”) are twin sisters. I had thought that Anna and Anna were the same person, as they are indiscernible. But Anna and Anna/Manuela are two people, and necessarily distinct entities, and hence I discover a posteriori the necessary fact that Anna is Manuela; and I discover a posteriori the necessary fact that Anna is not Anna! But I do not discover that I was contradicting myself when I thought that Anna was jogging in the park, even if it is impossible that Anna and Manuela were the same one.

(And I do not discover that whatever person with my favorite description of Anna was not whatever person with my favorite description of Anna.)

So we discover that Hesperus *is* Phosphorus, i.e., that “they” is the same entity, that “two” objects is in fact, and necessarily, the same one. Does the content of our thoughts before our discovery affect the kind of discovery we make when we discover that an identity statement is true? Perhaps it does, but not in the sense of being part of the discovery. On the other hand, the discovery is empirical, so it has cognitive value and propositional content. Its value and content is factual: we learned something about someone or about something; something about how the world is.

Not much ado about this. The relation of identity helps us to trace entities (including ourselves) in our world, in space and time, or even in our mind or our language. Every day, every hour, every instant ... we make discoveries of identity. You are the same you I saw yesterday. I see that Mary is still Mary today or there. This is ... same this. Constant discoveries and verifications of identity take place all the time. But they do not take place a priori. They are empirical facts that we ordinarily learn every day. Sometimes we make mistakes about them; sometimes they surprise us; the majority of the times they pass on without special interest. But all of them are necessary a posteriori facts.

### III

In truth, we cannot explain in ordinary language what we learned when we discovered that Hesperus is Phosphorus. We cannot, in proper ordinary language, say that we learned that these two objects “was” one and, necessarily, the same; or that this one entity “were,” in fact (and necessarily), two. This seems to be the case with many metaphysical truths and principles, and the reason probably is that they play a most basic role in our conceptual structure.<sup>13</sup> You cannot say in proper English, when trying to state the principle of Identity of Indiscernibles, that if any two entities share all their properties then they is just one. But of course you understand

the principle, you know what it would be to discover (empirically, or a priori) that it is true, or to discover that it is false. So its truth or falsehood cannot be a question of linguistic meaning. Many metaphysical theses need to be written in a formal language deliberately developed to display them. But when you understand them, you see that they can be denied without contradiction. Other principles of metaphysics can be expressed in ordinary language, like the thesis that any entity is necessarily identical to itself. But all metaphysical theses must be philosophically, not logically, not linguistically, defended. These are some samples:

- (N)  $(x) L(x = x)$
- (NI)  $(x)(y)(x = y) \rightarrow L(x = y)$
- (ND)  $(x)(y)(x \neq y) \rightarrow L(x \neq y)$
- (CLL)  $(x)(y)(P)((Px \leftrightarrow Py) \rightarrow (x = y))$
- (NE)  $L(x)(\exists y)(x = y)$
- (EN)  $(x)L(\exists y)(x = y)$

(N) states that any entity is necessarily identical to itself; (NI) that any identical entities are necessarily identical; (ND) that any distinct entities are necessarily distinct; (CLL) says that any entities whatsoever which have all their properties in common are the same one entity; (NE) that, necessarily, all entities exist; (EN) that all entities exist necessarily. And there are more theses like these whose nature is not logical, but of a metaphysical kind, like the Barcan formulae (which are directly related to the acceptance of (EN)), etc. All of them need philosophical defense, none of them is logically necessary, and before they are introduced as axioms into one's preferred logical system, the idea of logical contradiction cannot be brought into discussion.<sup>14</sup>

To see why these principles are not logically necessary has significant consequences for our understanding of the relation between metaphysics and epistemology. The idea that there are metaphysical theses, like the six above, that stand in need of philosophical defense implies that they stand in need of a priori defense; for if they belong to metaphysics, they must convey some kind of modality. But their defense, being a priori, cannot be supported by logic or linguistics alone. I have claimed that the necessity that Hesperus is Phosphorus is not directly related to logical identity or contradiction. I have also claimed that the discovery of the fact that Hesperus is, and necessarily is, Phosphorus does not relevantly include any conceptual truth of the kind that Hesperus is, say, the evening star, and thus we do not discover that the evening star is Phosphorus (and *consequently*, that Hesperus is Phosphorus). Hence the necessity that Hesperus is Phosphorus is not conceptual necessity, nor is it directly related to conceivability.<sup>15</sup> And yet, necessity (modality, in general) can only be established or justified a priori, because no instance of empirical knowledge can provide for modal truths.<sup>16</sup> Modality is a feature in the world but we do not discover it by merely observing nude particulars or particular facts. Even particular instantiations of modal a posteriori truths, like the discovery that Hesperus necessarily is Phosphorus, rest upon a priori modal principles, as I will show.

Necessary metaphysical truths are a priori established by intuition. Kripke also appeals to intuition at many points in his writings, and some work is being done ultimately on linguistic intuition in, so called, experimental philosophy.<sup>17</sup> But this is not the kind of intuition I will deal with, and I do not think that Kripke appeals to linguistic intuitions either. Rather, I will follow BonJour in his claim that “a priori justification occurs when the mind directly or intuitively sees or grasps or apprehends ... a necessary fact about the nature or structure of reality” (1998: 15–16).<sup>18</sup> Intuitions are insights, experiences that guide the activity of cognition. The subject experiences that a fact is necessary, and that everybody else should have the same feeling before the same fact. As with any other employment of our cognitive faculties, intuition also requires practice and training. We could wrong ourselves in its results; and yet, our feeling compels universal assent. All of us see that self-consciousness, even by momentary doubt, necessarily implies existence; that torture should not happen; that nothing bad can be beautiful; or that nothing can be other than itself. Philosophers need many years of study and hard work to really grasp or attempt any defense of theses like these. And they could be finally wrong in their conclusions, even when strongly feeling that, given the same facts, nobody should deny them. For these principles stand at the very bases of our knowledge and understanding of reality. They are pure apprehensions of reason that primarily concern fundamental philosophical categories of existence, of beauty, identity, the person, the good, and so on.

But the mind does not see without being in touch with reality. No human being can think without being in a world. No ultimate truth, not even truths of logic, can be understood by a mind devoid of senses. The case of identity is paradigmatic. Identity is a metaphysical relation. We know empirically, a posteriori, the necessary fact that Hesperus is Phosphorus; but knowledge of the necessity of this fact fully depends on knowledge of the necessity of identity itself, and this knowledge is a priori. We see a priori that it is impossible that any entity is other than itself. Many, or most, true descriptions of an entity could have been false descriptions of it. Most of its properties might not have belonged to it. But the entity could have not been another entity. And this we have not learned empirically. We have the philosophical intuition that no entity can be another. This is why any empirical discovery of a fact of identity is at the same time the discovery of a necessary relational fact. The discovery that Hesperus is necessarily Phosphorus is one of these discoveries, hence fitting the following – I would say, Kantian – schema:

- |      |                                      |                |
|------|--------------------------------------|----------------|
| (NI) | (x) (y) (x = y) → L (x = y)          | [a priori]     |
|      | Phosphorus is Hesperus               | [a posteriori] |
|      | Phosphorus is, necessarily, Hesperus | [a posteriori] |

The same schema applies to other cases of empirical discoveries of necessary facts. We see a priori that a species necessarily belongs to its genus. So when we learn, empirically, that cats are not demons, we learn empirically a necessary truth, but our knowledge of the modality of this truth rests upon a priori philosophical intuition of the necessary relation of membership that holds between any species and its

genus.<sup>19</sup> The modality in question is known a priori, and thus we see that empirical facts instantiating the categories in question are necessary, for they are instances of a priori metaphysical theses concerning such categories. In these cases, and perhaps in all cases, our different faculties do not work apart. For philosophical purposes, one can distinguish between the job done by reason and the job of our senses or experience. But, in fact, the process is not the realization of a deduction, as the schema above could wrongly suggest, but the working in team of reason and experience: when we learn empirically that Hesperus is Phosphorus, we also see that Hesperus cannot be but itself.

This is not to say that (NI) is an abstraction from such a fact as that Phosphorus is Hesperus, even if (NI) is, as if it were, written in that fact. There is no abstraction because intuition does not work by abstraction. Rather, we have an insight into a part of the fundamental structure of reality, a structure, as Strawson (1959) wrote, that does not readily display itself on the surface of language, but lies submerged.<sup>20</sup> This metaphysical structure belongs to the framework of experience, and its categories, like Identity, are necessary transcendental conditions for experience itself.

Thus, and to conclude, when we learn that Hesperus is Phosphorus, that *Londres* is London, that Tulio is Cicero, or that Mount Gaurisankar is not Mount Everest, what we learn, or discover, or what surprises us is, *literally*, that Hesperus is Phosphorus, that *Londres* is London, that Tulio is Cicero, or that Mount Gaurisankar is not Mount Everest. This kind of discovery is a basic part of our everyday learning. It is essential for our management and driving in the world. The particular cases I have dealt with along this paper are surprising precisely because we had been wrong before about them; what we had thought were two distinct entities, we later discovered to be (necessarily) the same one. Or, in the case of Everest (and Anna), what we had taken to be the same entity, we later discovered to be (necessarily) distinct ones. The discovery of these facts is more unsettling still because they are necessary. And the possibility of being wrong about them implies checking again pieces of knowledge where our most fundamental intuitions about the structure of reality are primarily applied. The necessity of these facts belongs in the discovered relations, but no necessity can be empirically discovered without an a priori insight into the necessary character involved in the category that is actually instantiated, and grasped. And it is by intuition, the “passive” activity of reason, that the nature and necessary character of identity (and diversity) is seen to fit within the basic structure of reality in a particular fact.

### **Acknowledgements**

This paper was written thanks to the financial help received from the projects FFI2013-47849-P and FFI2015-65953-P, sponsored by the Spanish Ministry of Education. I am very grateful to people at different meetings in the University of Granada, where I presented some of these ideas. I also thank three anonymous referees for this journal for all suggestions they made on the original manuscript. Special thanks to Fernando Martínez-Manrique for his comments.

## NOTES

1. Kripke aims at applying the theory of direct reference to other expressions, like names for natural kinds, indexicals, or demonstratives. For example, in case Hyde would say, “Oh, but I am Mr. Hyde; I have no idea who Dr. Jekyll is,” “I,” “Mr. Hyde” and “Dr. Jekyll” all three have the same referent, and, in consequence, they make the same semantic contribution to Hyde’s statement. Kripke (2008/2011) insists that, if they had different Fregean senses, under his own interpretation of Frege, this would have the undesirable consequence that, say, “Descartes says ‘I think’” and “Descartes says that he thinks,” would express different propositions (but they do not, according to Kripke). On the other side, “I think,” when said by Descartes, and “I think,” when said by Kripke, express different propositions (according, both, to Kripke’s Frege and to Kripke, but for different reasons). Nevertheless, even if I believe that Kripke’s theory, and most of what I will be saying in this paper, also applies to these cases, the importance of the particularities of this other kind of expressions demands a separate treatment. So I will keep the discussion within the realm of proper names.

2. Marcus rests her view on the thesis, also maintained by Kripke, that proper names directly refer to their objects. See Marcus (1961).

3. See also lecture IV in Kripke (1973/2013), especially pages 153ff.

4. Holmes is no other than a pretend-man in a fiction, and a fiction is a pretend story where proper names also pretend to name. So any statement about Holmes of the type that he smokes a pipe, works with Watson, and so on, does not express a proposition, but a pretend proposition. Kripke also says that “Holmes,” in a different use of language, has a referent: the fictional character that exists as an abstract entity. This abstract entity was created by the mind of Conan Doyle and, of course, it does not smoke, nor has any friends – although our language allows us to *elliptically* attribute to it these properties. This entity is irrelevant for the understanding of sentences like “Holmes does not exist” (see 1973/2011: 65).

5. Sure, we understand something when reading the novel, and we believe something when believing a myth. And what we understand can affect, and change, our lives. Novels, myths or fiction tales, make us feel and think, and see. They have meaning but express no proposition. And whatever it is that we understand when reading a novel or believing a myth, whatever we pretend at the time we recognize that the characters do not exist, our knowledge and pretence are not metalinguistic in kind. (Unless doing linguistics, we do not read a novel with a special consideration to the role of words and names in the sentences.)

6. In fact, Mount Gaurisankar and Mount Everest are not the same mountain; but the descriptions seem to be correct.

7. This is, in a shell, Jackson’s (1998) and Chalmers’ (2004) view that epistemic possibility, couched in terms of conceivability, finally explains metaphysical possibility. I have argued against this view in García-Encinas (2012).

8. What we publicly believe, or say to believe, is not any private stuff in our minds. This is a piece of Kripke’s externalism about mental content, and it explains his adherence to the disquotation principle, which is said to generate the puzzle in part. Nevertheless, the present discussion shows, I think, that this principle, and also the translation principle, are not essential to generate what seems to be a more general form of the puzzle: Just like Pierre agrees that *Londres* is pretty and denies that London is pretty, before discovering that Hesperus is Phosphorus, people agreed that Hesperus is the evening star, and denied that Phosphorus is the evening star. I will be back to these ideas in the next sections.

9. I think this is Lewis's (1981) view. Lewis seems to add that Kripke should not assume epistemic transparency ("knowledge of essences," he writes) at the time he (Kripke) defends a referentialist position. The path to reconciling contradiction and rationality, for solving or thinking about Kripke's puzzle, has also been followed by Kallestrup (2003), among others. Other philosophers (e.g., Sosa, 1996; Constantinescu, 2007) have centered the discussion on questions about translation, or on questions of belief ascription (e.g., Kemmerling 2006). However, my aim is not to propose a solution to this puzzle. Rather, I want to consider the common assumption that the puzzle conveys a contradiction, in the sense that Pierre *would discover that he was holding contradictory beliefs*, as a possible solution to the discovery case. For I think that Pierre's (plausible) discovery that London is *Londres* and the discovery that Hesperus is Phosphorus are, in essence, discoveries of the same kind; but, by themselves, none of these discoveries make us conscious of previous hidden contradictions: at most, they could make us reject some old beliefs, making us conscious of their falsity.

10. If you do not like this paraphrasing of the principle, just state it in terms of logical substitution.

11. Furthermore, it is not contradictory to deny the identity of Hesperus and Phosphorus, after it has been discovered, and generally proven, that Hesperus is, and necessarily is, Phosphorus. Likewise, it is not contradictory to deny that the sun rises, even when noting the fact that it rises every morning. That the sun does not rise is just plainly false, for it rises every morning. Reality does not contradict belief or language: it makes belief or language false or true. It seems contradictory to affirm that the sun rises and the sun does not rise (although, in truth, we should have first determined the formal system where this contradiction arises), but this is not the kind of affirmation we are dealing with. It is perfectly possible to deny a contingent truth: you say something false. In the case of a true statement of identity, its falsehood is impossible for it is a necessary truth; but the impossibility of this falsehood is not due to contradiction.

12. Reality does not contradict itself: not because it cannot fit into a logical system, but because the idea of contradiction does not apply. Contradiction applies only within the system.

13. This difficulty in writing or saying them in ordinary language could also be a source of the illusion of contradiction in cases like Pierre's. When you express some of these principles in ordinary language, either you feel that you are not really saying what they propose, or you are prone to grammatical mistakes as in "both is one."

14. A minimal logical rule of substitution plus, e.g., Modus Ponens. Substitution is a form of Leibniz's Law:  $(x) (y) ((x = y) \rightarrow (Px \rightarrow Py))$ , which you could accept as logically necessary. Pierre is faithful to this minimal logic. But you could also try to deny substitution and Modus Ponens – perhaps you have in mind the unorthodox philosophers Heidegger or Derrida, for a denial of the first; or the Tortoise of Lewis Carroll (1895) for the denial of the second. It would seem that this denial leads the way to the realm of plain logical contradiction. But even this is far from being obvious.

15. I have extendedly argued (García-Encinas, 2015) that neither analyticity nor the activity of conceptual thought and imagination can justify modal metaphysical principles.

16. The usual reason behind the idea that modal knowledge is a priori is that empirical experience is always subject to error: empirical experience can deceive us, so it cannot be the source of necessary knowledge. I think this reason is highly misleading. Error is part of human nature, just the same. And a priori means can be as fallible as any other. The reason why necessity can only be known a priori is that empirical knowledge gives you facts, but it



will not tell you whether those facts are contingent or not; it will not tell how they must, or should, or could be.

17. Machery et al. (2004) have run an empirical study where speakers of different languages were asked about the Kripkean case of Gödel and Schmidt. An analysis and defense of the main arguments that question the possibility of using this kind of intuitions for metaphysical purposes and, in particular, for the understanding of proper names can be read in Pérez Otero (2017).

18. Bealer (2002: 73) also thinks of intuition as a *sui generis* experience, an attitude that is the source of a priori necessary knowledge. And Chudnoff (2011) has made a nice case for the idea that intuitions are predoxastic experiences that represent abstract matters as being a certain way. Differences aside, their main point is that, when intuited, necessity is not inferred or deduced, but directly experienced.

19. Kripke (1980: 114n) says he can offer something like an a priori proof for the principle that: “[I]f a material object has its origin from a certain chunk of matter, it could have not had its origin in any other matter.” He did not really offer the proof but he did claim that only because we know a priori that original material constitution is necessary for certain kinds (material objects), we can discover a posteriori some properties of them that are necessary. Thus, when we discover a posteriori that a particular table is not made of ice, but of wood, we discover empirically a necessary property of the table – which is not to say that we discover empirically that the property is necessary (see Kripke, 1971: 88).

20. And then we can work in a language that expresses this knowledge, the language where (NI) is best written. (NI) is a metaphysical principle. But despite the general tendency to the contrary, it could be the case that even logical truths, or truths that are usually believed to be necessary but devoid of metaphysical content, like  $x = x$ , are finally grounded in metaphysics (and aesthetics).  $x = x$  could be a logical form of our intuition of Identity.

## REFERENCES

- Bealer, G. (2002). “Modal Epistemology and the Rationalist Renaissance,” in T. Gendler and J. Hawthorne (eds.), *Conceivability and Possibility*. Oxford: Oxford University Press, 71–125.
- Boghossian, P. (1994). “The Transparency of Mental Content,” *Philosophical Perspectives* 8: 33–50.
- Bonjour, L. (1998). *In Defense of Pure Reason*. Cambridge: Cambridge University Press.
- Carroll, L. (1895). “What the Tortoise Said to Achilles,” reprinted in D. Hofstadter (1979), *Gödel, Escher, Bach: An Eternal Golden Braid*. New York: Basic Books, 43–45.
- Chalmers, D. (2004). “Epistemic Two-Dimensional Semantics,” *Philosophical Studies* 118: 153–226.
- Chudnoff, E. (2011). “What Intuitions are Like,” *Philosophy and Phenomenological Research* 82(3): 625–654.
- Constantinescu, C. (2007). “Translating Kripke’s Pierre,” *The Reasoner* 1(7): 5–6.
- Jackson, F. (1998). *From Metaphysics to Ethics*. Oxford: Oxford University Press.
- García-Encinas, M.J. (2015). “The Role of Intuition in Metaphysics,” *Teorema* 34(3): 79–99.
- García-Encinas, M.J. (2012). “On Categories and A Posteriori Necessity: A Phenomenological Echo,” *Metaphilosophy* 43(1/2): 147–164.

- Kemmerling, A. (2006). “Kripke’s Principle of Disquotation and the Epistemology of Belief Ascription,” *Facta Philosophica* 8: 119–143.
- Kripke, S. (1971). “Identity and Necessity,” reprinted in Kripke (2011), 1–26.
- Kripke, S. (1972/1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Kripke, S. (1973). “Vacuous Names and Fictional Entities,” reprinted in Kripke (2011), 52–74.
- Kripke, S. (1973/2013). *Reference and Existence: The John Locke Lectures*. Oxford: Oxford University Press.
- Kripke, S. (1979). “A Puzzle about Belief,” reprinted in Kripke (2011), 125–161.
- Kripke, S. (2008). “Frege’s Theory of Sense and Reference,” reprinted in Kripke (2011), 254–291.
- Kripke, S. (2011). *Philosophical Troubles: Collected Papers I*. Oxford: Oxford University Press.
- Lewis, D. (1981). “What Puzzling Pierre Does Not Believe,” *Australasian Journal of Philosophy* 59(3): 283–289.
- Machery, E., R. Mallon, S. Nichols, and S. Stich (2004). “Semantics, Cross-Cultural Style,” *Cognition* 92(3): 1–12.
- Marcus, R.B. (1961). “Modalities and Intensional Languages,” *Synthese* 13(4): 303–322.
- Pérez Otero, M. (2017). “Teorías de la referencia, ‘filosofía experimental’ y calibración de intuiciones,” *Theoria* 32(1): 41–62.
- Sosa, D. (1996). “The Import of the Puzzle about Belief,” *The Philosophical Review* 105(3): 373–402.
- Strawson, P.F. (1959). *Individuals: An Essay in Descriptive Metaphysics*. London: Methuen.