Unification as a Measure of Natural Classification *

Victor GIJSBERS

Received: 07.02.2013
Final version: 14.07.2013

BIBLID [0495-4548 (2014) 29: 79; pp. 71-82]
DOI: 10.1387/theoria.7235

ABSTRACT: Recent interest in the idea that there can be scientific understanding without explanation lends new relevance to Duhem's notion of natural classification. According to Duhem, a classification that is natural teaches us something about nature without being explanatory. However, his conception of naturalness leaves much to be desired from the point of view of contemporary discussions. In this paper, I argue that we can measure the naturalness of classification by using an amended version of the notion of unification as defined by Schurz and Lambert. If this thesis is correct, it leads to a better conceptual understanding of scientific understanding, and also gives the nascent literature on this topic some much-needed precision.

Keywords: unification; understanding; explanation; natural classification; Pierre Duhem.

RESUMEN: El reciente interés en la idea de que puede haber comprensión científica sin explicación renueva la relevancia de la noción de clasificación natural propuesta por Pierre Duhem. Según este autor, una clasificación que es natural nos enseña algo sobre la naturaleza sin ser explicativa. Sin embargo, su concepción de "naturalidad" deja bastante que desear desde la perspectiva de las discusiones contemporáneas. En el artículo razono que podemos medir la naturalidad de una clasificación mediante una versión retocada de la noción de unificación, tal como es definida por Schurz y Lambert. Si dicha tesis es correcta, se logra un mejor entendimiento de la comprensión científica; y también se gana, para la naciente literatura sobre este tópico, una precisión sumamente deseable.

Palabras clave: unificación; explicación; clasificación natural; Pierre Duhem.

1. Introduction

In the past few years, philosophers of science have become interested in the topic of scientific understanding as separate from that of scientific explanation. Several authors argue that there may be understanding without explanation (e.g. De Regt 2001, 2009; De Regt & Dieks 2005; Lipton 2009; Gijsbers 2013). But the idea that we can gain understanding of nature without being able to explain the phenomena is not new: it can already be found in Pierre Duhem's seminal *The Aim and Structure of Physical Theory* (1991 [1914]). In that book, Duhem makes a distinction between explanation and classification. He then rejects explanation as an aim of science; but suggests that classifications, when they are *natural* classifications, can nevertheless give us insight into what is really going on in nature.

Even though his criticism of explanation is outdated given contemporary ideas about explanations, Duhem's conception of the distinction between explanation and natural classification is not. Natural classification is a good candidate for non-explanatory understanding, and as such is a topic of considerable interest for a

* I would like to thank Gerhard Schurz and Alexander Gebharter for inviting me to participate in the workshop that led to this paper, and Stathis Psillos for pointing me to Pierre Duhem.

philosopher of science. However, Duhem's own characterisation of naturalness leaves much to be desired. If we wish to use his ideas, we must come up with another conception of naturalness. Following a suggestion in Gijsbers (2013) that successful classification might be identified with unification, I want to explore the possibility that we already possess a measure of the naturalness of a classification: it is the notion of unification developed in Schurz and Lambert (1994).

In section 2, I will present Duhem's thinking about explanation and natural classification, and point out some problems with his conception. This will not only acquaint us with the intuitions underlying the idea of naturalness, but it will also allow us to formulate a list of desiderata for a measure of naturalness. I will then present a simplified and amended version of Schurz and Lambert's theory of unification (section 3), and I will argue that it matches the formulated desiderata (section 4).

This conclusion supports the idea that successful classification and unification can be identified with each other, which Gijsbers (2012) arrived at using very different reasoning. If correct, it would also mean that we can apply the technically sophisticated and quasi-quantitative tools that were developed by Schurz and Lambert in the context of unification to the topics of classification and understanding.

2. Duhem's notion of natural classification

The *locus classicus* of the idea that there can be scientifically important classification that gives understanding but not explanation is Pierre Duhem's *The Aim and Structure of Physical Theory* (1991 [1914]). At the beginning of his book, Duhem describes two different ways of thinking about physical theories. Some people believe that physical theories *explain* experimental laws, and that the aim of physical theory is explanation. Others believe that the aim of physical theory is to *summarize and classify* experimental laws. Duhem, then, starts with a sharp distinction between explanation and classification; and he goes on to argue that aiming for explanation is entirely mistaken.

We must note that Duhem's notion of explanation is much more metaphysical than that of contemporary philosophers. To explain, he writes, is "to strip reality of the appearances covering it like a veil, in order to see the bare reality itself" (p. 7). Explaining thus commits the explainer to an appearance-reality distinction, to the idea that we can get some kind of epistemic access to the bare reality behind the appearances, and to the idea that every explanation is a description of this bare reality.

Based on this metaphysical conception of explanation, Duhem argues that it cannot be an aim of science. For, he points out, whether something can be accepted as an explanation or not depends not just on our scientific discoveries, but even more on our basic metaphysical theories about the nature of bare reality. Duhem then sketches a bleak picture of the endless and undecidable battles between the metaphysical schools of the atomists, the Cartesians, the Newtonians and the Aristotelians. If we want our sciences to exhibit progress—and of course we do—we must not make them dependent on metaphysics. Metaphysics is merely a source of unending differences of opinion, a quagmire into which anyone unwise enough to seek explanations will surely sink.

This criticism of explanation has little bearing on contemporary discussions of the topic. Unlike Duhem, few philosophers now believe that all explanation consists in scientific facts being explained by underlying metaphysical principles. Most current theories of explanation (from Hempel's DN-model to Kitcher's unificationism and Woodward's interventionism) take as paradigm cases the explanation of events by other events, or of regularities by other regularities, two cases to which a metaphysical appearance-reality distinction seems entirely unimportant. Some theorists, like Salmon and Craver, do emphasize what we could describe as finding initially "hidden" mechanisms that are responsible for the "apparent" phenomena; but they believe that those mechanisms will be found by empirical science, not by metaphysical reflection. Our conception of explanation has changed since Duhem was writing; and we therefore will not wish to adopt his rejection of explanation.

But Duhem's distinction between explanation and classification remains interesting. So let us go on to discuss his notion of classification.

Duhem gives the following, oft-quoted, definition of a physical theory:

> It is a system of mathematical propositions, deduced from a small number of principles, which aim to represent as simply, as completely, and as exactly as possible a set of experimental laws. (p. 19)

One of the uses of physical theory so conceived is to aid the human mind in remembering and using all the knowledge of nature it has acquired (p. 21). But, Duhem argues, a physical theory is more than just an economical representation of the experimental results: it is also a 'classification' (p. 23). What is a classification?

In any physical domain, many experimental laws will be found. Physical theory groups some of these laws together; it gives us, "so to speak, the table of contents and the chapter headings under which the science to be studied will be methodically divided, and it indicates the laws which are to be arranged under each of these chapters" (p. 23-24). Using a different metaphor, Duhem likens a physical theory to a utility cabinet where all the tools used for a single task are grouped together. In that way, the physicist can easily find his tools, and can check that he has not forgotten to apply any of the suitable tools. For instance, if the physicist knows that he is studying a diffraction phenomenon, he will only have to look under that heading: he will not lose time trying to apply laws pertaining to refraction or reflection, and he will not forget to use any of the diffraction laws.

This is a classification. But a *successful* theory—and we will look at the criterion of success in a moment—persuades us that it is more than merely an arbitrary filing system; it persuades us that it is a *natural* classification. A natural classification is a classification where

> those ideal connections established by [the scientist's] reason among abstracted conceptions correspond to real relations among the associated [phenomena] brought together and embodied in his abstractions. (p. 25)

For instance, a zoologist classifies the vertebrates. If the grouping he arrives at—say, a grouping in which the whales are closer to the hippopotami than to the fishes—corresponds to a real relation among the actual animals, this classification is a natural classification. But, and this is crucial for the difference that Duhem makes between explanation and classification, the zoologist doesn't have to know what this real relation is; he doesn't even have to have any idea or hypothesis about it; and if he does have an idea about it, it doesn't matter if it is wrong. Considering the idea that the real relation underlying the classification of vertebrates is that of closeness in the evolutionary tree, Duhem writes:

> And if physiology and paleontology should someday demonstrate to [the zoologist] that the relationship imagined by him cannot be, that the evolutionist hypothesis is controverted, he would continue to believe that the plan drawn by his classification depicts real relations among animals; he would admit being deceived about the nature of these relations but not about their existence. (p. 25)

Whether we accept a classification as natural is therefore not dependent on our knowing which real relations are captured by the ideal relations of our theory. The understanding we gain from classification can be integrated into an explanatory theory of reality, but can also exist alone, and it will survive the rejection of any and even all explanations.

A natural classification, then, is a classification that captures the reality of nature without making explanatory claims. But why would we ever accept a classification as natural, given that, according to Duhem, we do not have access to the real relations themselves? It is the *success* of the classification that makes scientists believe it is natural, that its logical order reflects the ontological order of nature (p. 26).

What criterion of success does Duhem have in mind? Minimally, the classification must be an economical system for classifying the experimental laws. But there is another kind of success that, according to Duhem, makes us especially likely to believe that the classification is natural: *predictive success*. A sufficiently rich theory will allow us to deduce new experimental consequences that are not consequences of the hitherto established experimental laws. These consequences can be tested. Using an argument that strongly resembles the no miracle argument for scientific realism (see for instance Psillos 1999, p. 70ff), Duhem claims that if a classification is merely an arbitrary system of categorisation, such predictions are highly unlikely to be true; but that if the classification corresponds to the real structure of the world, they are in fact quite likely to be true. Successful predictions based on the theory are thus particularly strong motivations for believing that we have indeed found a natural classification.

This combination of a dismissal of explanatory theories and an acceptance of what can be read as an argument for scientific realism has led to quite a bit of literature about Duhem's position in the realism debate (e.g., Lugg 1990; Darling 2003; Dion 2013). In addition, his ideas about how a natural classification allows us to grasp the *structure* of a relationship but not the *nature* of the underlying reality has been an important source of inspiration for structural realism (Worrall 1989). But our concerns in this paper are not with realism and antirealism, so I will bracket these issues.

Can we use Duhem's notion of a natural classification in contemporary discussions about scientific understanding? The basic idea—that there is a kind of classification that gives us understanding without explanation, and that we recognize this kind of classification from its inductive success—seems very usable. But Duhem's own account of naturalness has several weaknesses that we will have to overcome.

Duhem defines a natural classification as a classification in which the ideal relations postulated by the theory correspond to the real relations in nature. For our purposes, there are three problems with this approach:

- Duhemian naturalness is a primarily metaphysical concept. But we would like to connect naturalness to understanding, which is a primarily epistemic concept.
- It is unclear on Duhem's terms whether we can ever be justified in believing that a classification is natural; and clarifying this would mean taking up a position in the realism debate. When we are theorising about scientific understanding, we would prefer to have a measure of naturalness that can be uncontroversially applied; and we would prefer to remain agnostic about realism and antirealism.
- Duhem's notion of a 'real relation' is anyway unclear. Given a set of objects, there is presumably always *some* relation that they bear to each other. How do we select the 'real' relations from among all relations? And can we do this without sinking into the quagmire of metaphysics?

Next to the criterion that the ideal relations must correspond to the real relations, the second ingredient of Duhem's approach is the idea that we can recognize naturalness from inductive success. We will assume that Duhem is onto something here: a classification is natural, and gives us understanding about nature, if it describes more of nature than we explicitly put into it. Given the problems with Duhem's definition of naturalness in terms of correspondence with real relations, it seems worthwhile to investigate whether it is possible to use inductive success as a *definition* of naturalness rather than merely a *sign* of it.

However, Duhem's own notion of inductive success is vague. Furthermore, where he becomes more concrete—in putting special emphasis on unexpected successful predictions—the results seem to be counter-intuitive. For if a theory has more inductive success if its correct predictions were unexpected, then its naturalness will depend on the idiosyncratic details of the history of science. But naturalness and understanding (perhaps unlike justification) seem to be notions that should not have that dependence, as we know from discussions about the problem of old and new evidence. We want a notion of inductive success that is independent of the development of science.

This brings me to the following set of desiderata for a measure of the naturalness of classifications:

1. It should allow us to actually measure naturalness (e.g., no dependence on an unknowable hidden reality).
2. It should be a measure of inductive success.
3. That inductive success should be conceived of ahistorically.
4. It should be applicable to classifications broadly conceived: not just 'grouping objects under headings', but also complicated scientific theories.
5. When we look at examples of clearly natural or clearly unnatural classifications, it should give the right verdict.

Of this list, 1, 2 and 3 have already been discussed, and 5 is self-evident. So let me just say a few words about 4. For Duhem, classification is a very broad notion: optics is a classification of phenomena, and so is Newtonian physics, even though these theories do much more than just dividing the phenomena into a set of mutually exclusive and jointly exhaustive sets (which would be a more narrow definition of classification). Since such theories presumably give us scientific understanding—even if their explanatory, e.g., causal, claims would turn out to be false—we will join Duhem in thinking of them as potentially natural classifications. This means that our measure of naturalness must be applicable to such theories.

In sections 3 and 4 I will argue that Schurz and Lambert's theory of unification fits these desiderata, and is therefore a good candidate for a measure of naturalness.

3. Schurz-and-Lambert-unification

In this section, I will first present a simplified and amended version of the theory of Schurz and Lambert 1994 (see also Schurz 1999), and then explain what the simplifications and amendments consist in.

We start with the cognitive corpus *C* of a scientist or a scientific community. *C* consists of two parts: the set of accepted statements *KNOW*, and the set of accepted inference patterns *I*. We will not pay much attention to *I*, assuming that it contains just the generally accepted forms of deductive, inductive and probabilistic reasoning, and no 'weird' or 'deviant' forms of reasoning. We will also assume that *KNOW* is logically closed under the accepted inference patterns—i.e., we assume a situation of logical omniscience.

KNOW is of course infinite. Because we want to count statements later, we need to use the method of knowledge representation by relevant elements (Schurz and Lambert 1994, 88-91) to pair *KNOW* down to a finite set of relevant elements—the details of this procedure will not be described here (but see Schurz's discussion in this volume, pp. 60-61). We call this paired down set *K*.

We next make a distinction in *K* between *data* and *hypotheses*. A datum is a statement describing a natural phenomenon that we have observed, while a hypothesis is a statement that is part of the theory we have formulated to summarize and classify (and maybe explain) the data. As used here, this distinction is not supposed to commit us to the idea of a pure observation language in which the data are 'given'. Rather, what we need is a distinction—made in any way that one deems philosophically acceptable—between that which needs to be classified and that which classifies. Perhaps the idea of data can be clarified using the notion of an observation language; perhaps it can be clarified using more sophisticated ideas, such as Azzouni's (2000) notion of evidentially central content; or perhaps we have to think of data as the set of statements that a scientific community accepts as beyond rational discussion. For our purposes we can remain agnostic about this. We will also ignore the question whether what are data will change when theory changes, interesting though this question is. Only two things are important: that the distinction between data and hypotheses can be made; and that the distinction has enough pre-theoretical clarity that we can discuss examples of classification without having to make the distinction fully explicit. That pre-theoretical clarity does seem to exist. For instance, everyone will agree that "this lump is yellow" is a better candidate for a datum than "the atomic number of gold is 79".

With the distinction between data and hypotheses in place, we can start defining unification. The set K is to be divided into two mutually exclusive and jointly exhaustive subsets: the *basic phenomena* and the *assimilated phenomena*. This division must follow the rule that all assimilated phenomena can be inferred from the basic phenomena using inferences in I.

That criterion will usually not pick out a unique division. The second criterion is that of all divisions that conform to the first criterion, the division that achieves the *greatest unification* is the division that must be chosen.

How do we calculate which division leads to the greatest unification? We use the following four rules:

- A *datum* that is *basic* gives 0 unification.
- A *datum* that is *assimilated* gives $+W_d$ unification.
- A *hypothesis* that is *basic* gives $-W_h$ unification.
- A *hypothesis* that is *assimilated* gives 0 unification.

The rationale behind these rules is as follows. Basic data are our starting point, and so we assign them neither positive nor negative unification. Our knowledge doesn't become more or less unified just because we make more observations; unification comes from our classifications.

When we introduce hypotheses, these come with an intrinsic unification cost. Hypotheses that do not allow us to assimilate any data are just dead weight, wheels on the machine that do not turn; and therefore we assign them a negative value $-W_h$. But hypotheses can pay for themselves by assimilating data. Assimilated data, data that can be derived from our classificatory theories, are the positive currency of unification, and we therefore assign them a positive value $+W_d$.

Finally, hypotheses that can be inferred from other hypotheses are assigned a zero value: they do no increase unification, but neither do they needlessly add to the complexity of the system. (Once we have paid for Newton's laws, we get Kepler's laws for free.)

To illustrate this, let us look at a simple example. Suppose we have n lumps of matter, and suppose that K consists of the following $3n$ statements, with i running from 1 to n:

> Lump i is yellow. Lump i has a melting point of 1337 K. Lump i is quite soft.

None of these statements can be inferred from any of the others; and we assume that all of them are data. So we have $3n$ basic data, which gives us a total unification of 0.

Now we add the following statements to K:

> Lump i is gold. All gold is yellow. All gold is quite soft. All gold has a melting point of 1337 K.

None of these statements are inferable from any of the others, so we here have $n+3$ basic hypotheses. However, the $3n$ basic data statements we started out with are all inferable from these hypotheses, so they now count as assimilated data. The total unification of our new K is therefore $3nW_d - (n+3)W_h$.

Whether this is positive or negative depends on the values of W_d and W_h. Since a hypothesis must surely assimilate more than one datum before we are justified in adopting it, $W_h > W_d$. But does a hypothesis have to assimilate 2, or 3, or 5, or 11 data before we are justified in adopting it? There is probably no precise answer to such a question. Furthermore, the assumption that all data and all hypotheses come with the same values W_d and W_h should be regarded as a simplification. Schurz and Lambert defend the plausible claim that general hypotheses like "all gold is yellow" are more costly than particular hypotheses like "lump 118 is gold". Again, there is probably no way to determine precisely how much greater—different scientists will have different opinions. This means that calculating unification is not an exact science; but neither does it have to be. Scientists themselves can have rational disagreements about whether or not a certain classification is worth the trouble, or which of several classifications is the most natural one (as the many debates about biological classification have amply illustrated).

But we can say this: given enough lumps and enough properties that can be linked to being gold, there will certainly be a point at which the introduction of a theory about gold will increase unification, and will therefore be worth the trouble. This is exactly what we would expect.

In section 4, I will argue that SL-unification (as I will abbreviate 'Schurz-and-Lambert-unification') is a good measure of the naturalness of a classification. But I first need to point out several differences between my presentation of Schurz and Lambert theory.

In condensing their 56-page paper to two pages, I have obviously made many simplifications. The most interesting of these is that Schurz and Lambert's distinguish not just between basic and assimilated statements, but that they also add the categories of dissimilated and heuristically assimilated statements. This is a fruitful and important idea, but too involved to explain here. I have also left out the erotetic framework adopted by these authors; their interest in shifts in opinion; and the possibility of changing the store of accepted inferences. The reader is advised to consult the original article for its many fascinating ideas.

But there is something else that I have left out not because of reasons of space, but because for our current purposes, it must be removed from the notion of SL-unification. Schurz and Lambert want to ensure that the inferences of the most unifying theory correspond to *real relations* between the phenomena. As an important example, they prefer inferences that capture the real causal structure of the world to inferences that do not. Because of this preference, they give a unification bonus to inferences that fit into a system of general principles of causality.

Now causal inferences are typically explanatory, as the dominance of causal theories of explanation testifies, and a preference for causal inferences is no doubt due to Schurz and Lambert's wish to give a unificationist theory of explanation. But since our purpose is to look at classifications prior to the addition of any explanatory principles (including causation), we should not adopt this idea.

There are additional reasons for preferring a notion of SL-unification that remains agnostic about explanatory principles. Gijsbers (2007) argues that unification cannot be a criterion for explanation; if this is correct, then adding a principle to our notion of unification based on the idea that unification can be such a criterion is misguided. Furthermore, Gijsbers (2013) argues that unification leads to a kind of understanding that is simply different from the understanding provided by explanation. So while it may be true that scientists prefer, say, causal theories to non-causal ones, it seems a good idea to keep this preference conceptually distinct from their preference for unified theories. For these reasons, I believe that we should strip away the preference for causal theories from the notion of SL-unification.

4. Unification and naturalness

We are now in a position to consider the following thesis:

> SL-unification is the measure for the naturalness of a classification. If it makes sense to speak of *the* natural classification of a domain of phenomena, that natural classification will be the theoretical structure that achieves the highest amount of SL-unification.

Since we already formulated five desiderata for a measure of naturalness, we can simply check whether SL-unification has these desired properties.

1. SL-unification can indeed be used to actually measure naturalness: it is calculated based on the set of accepted statements, not based on anything that is unknown or even unknowable.
2. SL-unification is a measure of inductive success. Positive unification is generated by inferring data from other statements; if we can predict more of the data, we have higher unification—assuming, that is, that we do not have to add a disproportionate amount of new theory in order to increase the amount of data that can be inferred. This is as it should be: we reach maximal inductive success at the point where we have the best balance between the amount of data we can assimilate and the amount of theory we need in order to do that.
3. SL-unification is calculated based on the set of accepted statements, independent of how and in what order those statements came to be accepted. It is thus appropriately ahistorical.
4. SL-unification can be calculated for any theory that has inferential consequences, which is, presumably, anything that could be called a theory. It can therefore be applied to the entire range of scientific theories.
5. Whether SL-unification agrees with our intuitive judgements about naturalness is an open question that cannot be decided in a simple way.

We see that SL-unification performs well on the first four desiderata. To get more insight into the fifth, we will look at two examples: a simple case also discussed by Schurz and Lambert under the heading of 'spurious classification of the first kind', and a more complicated case involving double classification.

Example 1: virtus dormitiva

Suppose we have a set of *n* substances that put people to sleep. The *n* statements "substance *X* puts people to sleep" are basic data. We now add the *n* hypotheses that "substance *X* has a virtus dormitiva", and the hypothesis "anything with a virtus dormitiva puts people to sleep". This is a well-known example of an unnatural classification; so it should have negative SL-unification. Does it?

The positive unification is $n\ W_d$. The negative unification is $(n+1)\ W_b$. Since $W_b > W_d$, we get the correct answer that the total SL-unification will be negative. Adding new theoretical categories that correspond 1-on-1 with pre-existing phenomenal categories does not lead to natural classifications.

Example 2: whales as mammals and as fish

Suppose that we already have theories of mammals and fish in place (they have been paid for, and we will not consider their costs and general benefits in the rest of example). For simplicity, we assume that these theories have the following form: if we have a statement that says that a certain animal is a mammal (or a fish), we then have non-deductive inferences in *I* that allow us to infer that this animal has each of a certain set of properties. So, if we know that cats are mammals, we can infer that cats have two eyes, have a backbone, are hot-blooded, and so on. The inferences are non-deductive because we allow for the possibility that they are blocked by other accepted statements, e.g. the statement "in terms of laying eggs, the platypus is not mammal-like".

Suppose furthermore that we have already classified whales as mammals, because many of their properties get assimilated once we do so. Here is the question we need to ask: does it make sense to *also* classify whales as fish? After all, several of their properties did not get assimilated by classifying them as mammals, but would get assimilated by classifying them as fish. The answer should be no: classifying whales as fish has been considered an unnatural move by modern biologists. Does the notion of SL-unification support this judgement?

Take all properties that our theory of fish ascribes to fish. This set can be divided into three parts. *X* contains those properties that whales have, but that already got assimilated when whales were classified as mammals. *Y* contains those properties that whales have, and that did not get assimilated when they were classified as mammals; and *Z* contains those properties that whales do not have. To give some examples, *X* contains 'have a backbone' and 'are warm-blooded'; *Y* contains 'have fins' and 'live in water'; and *Z* contains 'are cold-blooded' and 'have gills'.

What is the unification effect of classifying whales as fish? The properties in *X* do nothing. The properties in *Y* give a positive unification effect, since every statement of the form "whales possess property *a*" with *a* in *Y* gets assimilated. But the properties in *Z* give a negative unification effect, because we must block the incorrect inferences they give rise to. (In fact, the formally correct way to represent this negative unification effect is to move the statements that whales do not have the properties in *Z* to the category of dissimilated statements. But, as indicated in section 3, I have not discussed that part of Schurz and Lambert's theory. It does not matter for our example.)

Now Z is bigger than Y: once we control for the properties of mammals, whales share only a few properties with fish, while they lack many distinctively fishy properties. So the unification effect of classifying whales as fish will be negative. The notion of SL-unification allows us to correctly identify this attempted classification as unnatural.

(One might argue that in current biology, the classification of whales as fish would be rejected because it doesn't fit our ideas of evolutionary history. This may be the case; explanatory theories will often influence the way we classify things, and it is in no way the aim of this paper to argue that scientists often or generally rely on pure classification rather than on explanatory theories. But they *could* do so and still get some understanding of the natural world.)

Examples could and should be multiplied, of course, in order to strengthen the case that SL-unification captures naturalness. But in the limited space of this paper, these two examples will have to suffice. We thus see that SL-unification performs well on all five desiderata that we formulated for a measure of naturalness.

5. Conclusion

Recent interest in the idea that there can be scientific understanding without explanation lends new relevance to Duhem's notion of natural classification, since he argued that natural classifications can teach us something about nature without being explanatory. However, Duhem's conception of naturalness leaves much to be desired from our point of view. In this paper, I have argued that unification as defined by Schurz and Lambert has the desired properties for a measure of the naturalness of classification. If this is true, SL-unification measures a type of scientific understanding that until now has been somewhat neglected in the philosophical literature. It also means that we can use this notion to add some much-needed precision to discussions of scientific understanding.

Whether the identification between SL-unification and naturalness is tenable should of course be studied further; the few examples discussed in this paper cannot conclusively justify this identification. But we have at least some initial justification for believing the thesis.

Incidentally, if the thesis proves to be correct, this could also have effects on debates about structural realism. Since structural realists have often appealed to Duhem's ideas about classification, it might well turn out that the quasi-quantitative approach of Schurz and Lambert can be applied to case studies in the literature on structural realism. But judging the plausibility of such an approach lies beyond the scope of this paper.

REFERENCES

Azzouni, Jody. 2000. *Knowledge and Reference in Empirical Science*. London: Routledge.

Darling, Karen Merikangas. 2003. Motivational Realism: The Natural Classification for Pierre Duhem. *Philosophy of Science* 70: 1125-1136.

De Regt, Henk W. 2001. Spacetime Visualisation and the Intelligibility of Physical Theories. *Studies in History and Philosophy of Modern Physics 32*: 243-265.
—. 2009. Understanding and Scientific Explanation. In *Scientific Understanding: Philosophical Perspectives*, eds. Henk De Regt, Sabina Leonelli and Kai Eigner , 21-42. Pittsburgh: University of Pittsburgh Press.
—. and Dennis Dieks. 2005. A Contextual Approach to Scientific Understanding. *Synthese* 144*:* 137-170.
Dion, Sonia Maria. 2013. Pierre Duhem and the inconsistency between instrumentalism and natural classification. *Studies in History and Philosophy of Science* 44: 12-19.
Duhem, Pierre. 1991 (1914). *The Aim and Structure of Physical Theory*. Princeton, NJ: Princeton University Press.
Gijsbers, Victor. 2007. Why Unification Is Neither Necessary nor Sufficient for Explanation. *Philosophy of Science* 74: 481-500.
—. 2013. Understanding, explanation, and unification. *Studies in History and Philosophy of Science Part A* 44: 516-522.
Kitcher, Philip. 1981. Explanatory Unification. *Philosophy of Science* 48: 507-531.
—. 1985. Two Approaches to Explanation. *Journal of Philosophy* 82: 632-639.
—. 1989. Explanatory Unification and the Causal Structure of the World. In *Scientific Explanation*, eds. Philip Kitcher and Wesley C. Salmon, 410-505. Minneapolis: University of Minnesota Press.
Lipton, Peter. 2009. Understanding Without Explanation. In *Scientific Understanding: Philosophical Perspectives*, eds. Henk De Regt, Sabina Leonelli and Kai Eigner , 43-63. Pittsburgh: University of Pittsburgh Press.
Lugg, Andrew. 1990. Pierre Duhem's Conception of Natural Classification. *Synthese* 83: 409-420.
Psillos, Stathis. 1999. *Scientific Realism: how science tracks truth*. London: Routledge.
Schurz, Gerhard. 1999. Explanation as Unification. *Synthese* 120: 95-114.
—. and Karel Lambert. 1994. Outline of a Theory of Scientific Understanding. *Synthese* 101: 65-120.
Worrall, John. 1989. Structural realism: The best of both worlds? *Dialectica* 43: 99-124.

VICTOR GIJSBERS is assistant professor of philosophy at Leiden University. His research focuses on general theories of explanation; the concepts of unification and understanding; and the relation between causation and agency.

ADDRESS: Universiteit Leiden, Instituut voor Wijsbegeerte, Postbus 9515, 2300 RA Leiden, Netherlands. E-mail: v.gijsbers@hum.leidenuniv.nl