From *Water* to H_2O – What Reduction is About

Raphael van Riel
(Raphael.vanRiel@rub.de)

Abstract: In what follows I shall argue that an important notion of reduction depends on a four-place relation holding between expressions, concepts, properties, and events or states of affairs. I define this notion and argue against alternative accounts that are based on syntactic features of theories. Whilst these latter attempts fail to deliver a satisfactory explanation of why a certain theory or a certain expression reduces to another, the former can give a complete explanation of why, say, 'human pain' reduces to 'C-fiber stimulation' (if it reduces at all) or why the mind reduces to the physical. I briefly sketch the difference between the semantic approach that I favor, which is based on a particular notion of hyper-intensions, and classical model-theoretic versions of reduction.

Preliminary

In this paper, I shall discuss a particular concept of reduction, namely the concept of reduction that seems to underlie the following prominent claim: *The mind reduces to the physical.* What exactly this thesis states is a debated issue. The pertinent notion of reduction is often illustrated by way of examples. Here is one of them:

- Human pain reduces to C-fibre-stimulation.

Another example for this sort of reduction can be found in Fodor (1981, p. 150) and Kim (1993, p. 333):

- Water reduces to H_2O.

I shall use these examples as paradigmatic cases of reductionist claims, and I shall take their truth for granted.
Taking these examples as paradigmatic cases for reductionist claims, I restrict my use of the term 'reduction' to the way it is used in these cases. Other notions will be taken into account only if they seem to be capable of illuminating the notion of reduction I am interested in here. '*The* reduces-to-relation' refers to the relation of reduction corresponding to this notion (henceforth: 'reduction-relation')´.
It will turn out that the notion of reduction is primarily a notion of a certain sort of explanation-improvement. This is a familiar point. Nevertheless, many theories according to which the reduction-relation is an explanation-relation face the problem of trying to spell out the difference between reduced and reducing level or theory either in terms of syntactic relations or in terms of properties or phenomena, thus neglecting the fact that explanations are sensitive to conceptual contents of the sentences being connected by the binary-connective 'because'. I shall argue that reduction is primarily concerned with levels of descriptions and that levels of descriptions are levels of conceptual contents, rather than levels of properties, phenomena, events, or states of affairs. It will turn out that the reduction-relation is a four-place relation holding between expressions, conceptual contents, properties and events or states of affairs.

Philosophers have investigated notions of reduction from different angles. Some are interested in formal relations holding between theories, some are interested in criteria for an appropriate formalization of historical theory-succession. Others hope to define an ontological hierarchy according to which our world is organized using the notion of reduction. These latter philosophers often start with an intuitively appealing slogan, such as "If Xs are reduced, or reducible, to Ys, there are no Xs *over and above* Ys" (*cf.* Kim 2006, p 275 f. For a comment on this slogan, see below, footnote 6), and then try to give a more formal account of what it states. Despite these different interests, philosophers of mind, who belong to the camp of the latter, sometimes refer to formal approaches to reduction in order to explicate their notion of reduction (*cf.* Kim 1993, Crane 2001 (see below)).

In what follows, I shall focus on the most influential approaches to theory-reduction – syntactic and certain semantic approaches – and try to answer the following question: Can these approaches help to clarify the notion of reduction we are interested in here? As we shall see, these approaches face a serious problem: They are incapable of explaining what reduction consists in. This is partly due to the fact that they do not furnish us with the equipment we need to pick out the relevant entities – that is: conceptual contents. [1]

The paper is structured as follows: Firstly, I will argue that syntax-based accounts of reduction lack any explanatory power concerning the question what reduction consists in (Part 1). I will then argue that many theories reduce to other theories in virtue of relations holding between conceptual contents that are expressed by a theory's elements, properties and events or states of affairs (Part 2). In order to do so, I define a notion of synonymy that enables us to individuate conceptual contents. In the last section, I will compare the account presented below to classical semantic (model-theoretic) views of intertheoretic reduction and briefly explain why these latter accounts fail (Part 3). I finish by sketching an alternative way of modelling a theory's semantics, delivering individuation-criteria for conceptual contents or *hyper-intensions* that are based on set-theoretical notions.

Let me mention that I do not aim at defining a historical notion of reduction which seems to underlie many versions of intra-level reduction. The

[1] Since the account I argue for in this paper is, in a sense, a semantic account, too, I distinguish between classical semantic approaches and non-classical approaches.

examples I discuss are standard examples for reduction taken from the Philosophy of Mind debates. Hence, we will primarily be concerned with reduction of expressions of (or: explanations formulated in) ordinary language (e.g. folk psychology) to expressions of (or: explanations formulated in) scientific language, and with reduction of the corresponding theories.

1. Failing to answer the why-question

Ernest Nagel developed an influential model of inter-theoretic reduction, describing the goal of such reduction as follows:

> Reduction … is the explanation of a theory or a set of experimental laws established in one area of inquiry, by a theory usually though not invariably formulated for some other domain. (Nagel 1961, p. 338)

According to Nagel, a theory R (reduced theory) reduces to a theory B (base theory) iff there are some bridge-principles C, delivering criteria to map expressions of A onto expressions of B, such that B and C together (syntactically) entail A (ibid.). Note that this is a very brief version that is in need of a substantial refinement: According to the above definition, every theory (trivially) reduces to itself and every theory reduces to every necessarily false theory. But this sketch will suffice to give an idea of what syntactic theory reduction consists in. Even though this conception has been attacked for several reasons that I am not discussing here (*cf.* Putnam 1975, Fodor 1974), many contemporary reductionists develop their models in Nagel's spirit (*cf.* Schaffner 1993, Hooker 1981, Bickle 1998). These attempts have in common that they (i) take theories to be (derivatively) syntactically structured entities, i.e., sets of sentences, and, therefore, (ii) take theory reduction to be concerned with relations that are based on syntactic properties. Furthermore, if the definitions delivered by these accounts are supposed to illuminate what reduction consists in, it must be possible to *explain why* certain theories reduce to other theories referring to these definitions. The main problem is that these approaches to reduction fail to meet this criterion, i.e.: They are incapable of *explaining* what reduction consists in.

My argument is based on a simple observation: According to every non-semantic version of inter-theoretic reduction, the question of why a theory *T* reduces to another theory *T** is answered by referring to syntactic properties. Consider the following passage in which Jerry Fodor sums up what he takes to be the core idea of Nagel-style reduction:

> [T}heories in whose laws the expression 'water' (or its cognates) occurs will reduce to chemistry only if (a') chemistry contains some expressions other than 'water' (say 'H2O') such that (b') '(x) (x is water if x is H2O)' expresses a law. (Fodor 1981, p. 150)

Now, let us turn to the why-question. We can ask: Why does *T* reduce to *T**? All that the Nagelian theorist can do in reply is to point to syntactic features of *T*, *T**, and, according to some versions of inter-theoretic reduction, to bridge principles. But the syntactic features are not of interest in this context. To give an example: Let us assume that (human) pain is identical to C-fibre stimulation. If I ask why pain-science reduces to C-fibre-stimulation-science and you tell me something about the syntactic properties of 'pain', 'C-fibre stimulation' and general claims of the form 'For every F, F is G (and, maybe, *vice versa*)', I would be rather astonished. Putting the question this way, we do *not* want to know anything about syntactic entities like 'pain' or about properties these entities have in virtue of having their syntactic properties – which is obviously all that can be explained by referring to the entity's syntactic properties. Thus, the objection is that syntactic deduction is explanatorily irrelevant for understanding reduction (in the sense of `reduction´ we are interested in here). Therefore, syntax-based approaches to the notion of reduction cannot deliver illuminating definitions. Therefore, Nagel's intuition that reduction is concerned with the *explanation* of one theory by another, and that this sort of explanation has got something to do with syntactic deduction, is not compatible with the notion of reduction underlying our paradigmatic cases.

The notion of a bridge-principle in fact goes beyond considerations of mere syntactic reduction. Bridge-principles seem to be able to furnish us with the relevant information concerning the facts in virtue of which a certain theory reduces to another theory. But they do not do so in virtue of their syntactic structure, but rather in virtue of their semantics. Semantic accounts of the notion of reduction (including the version I shall outline

below) can be read as promoting criteria, which bridge-principles have to meet. If so, bridge-principles tell us something about the semantic values of 'human pain' and 'C-fibre-stimulation', and about relations holding between these semantic values. If this is correct, then syntactic entities reduce only derivatively in the sense that they reduce in virtue of their semantics. Consequently, no satisfactory explanation of why a theory reduces to another can be gained from purely syntactic accounts.[2]

Nevertheless, the idea that reduction is concerned with syntactic entities has, at least *prima facie*, its virtues: It seems to perfectly match the observation that if A reduces to B, then 'A' is to be replaced by 'B', at least in some contexts. But the fact that we cannot explain why A reduces to B on the basis of syntactic properties alone seems to show that expression-replacement can be subsumed under the concept of reduction only in a derivative sense.

What, then, is reduction primarily concerned with, if not with syntactic entities? Is it concerned with the objects of the theory? This is what Jaegwon Kim seems to suggest when introducing his model of functional reduction (*cf.* Kim 1993). In this case, the bridge-principle(-analogue)s tell us that a certain species or system bound (human…) kind (pain) is functionally equivalent to – and, according to a certain view of property- or kind-individuation, thereby identical to – a kind being described in a different vocabulary ('C-fibre stimulation'). Yet, identity and functional equivalence are *symmetric* relations, whilst the reduction-relation is asymmetric. Furthermore, when the project of reduction (i.e.: local reduction) really starts and the considerations about property-individuation, functional roles and conceptual analysis are put aside, Kim's account is to be understood in the Nagelian way (*cf.* Kim 1993, p. 328).[3]

2. An alternative view

In order to establish an alternative view, let me start from the notion of a *cognate* mentioned in the Fodor-citation above and state two observations. Similar to the aforementioned problem concerning the explanatory power of syntactic properties in contexts concerning reduction, we have to answer

[2] For a similar, though not identical worry see (Moulines 1984, p.55).

[3] For a detailed discussion of whether or not Kim's version of reduction is Nagelian in nature, see Marras (2002&2006).

the question of why cognates of 'water' should be replaced by 'H_2O' in the pertinent case.

The notion of a cognate is a semantic notion. I take Fodor to mean that 'water' and its French counterpart 'eau' are cognates. According to another reading, 'water' and 'iron' on the one hand and 'H_2O' and 'Fe' on the other count as cognates. Which reading Fodor actually intended is not necessary to establish; both aspects are interesting, I will now focus on the first one. Obviously, no syntactic property will deliver an illuminating criterion for what makes an expression a cognate of another expression. Nor will an ontological claim about (the property of being/the substance/the natural kind) water and (the property of being/the substance/the natural kind) H_2O do. After all, the property of being water is identical to the property of being H_2O – at least this is what the reductionist should believe. If these properties are identical, then the difference between *water* and *H_2O* cannot be explained by referring to the property that both terms can be used to ascribe. To repeat: Functional equivalence and identity are symmetric relations, whilst the reduction-relation is asymmetric. This observation has forced some philosophers to distinguish between ontological and explanatory aspects of reduction, or between two sorts of reduction. Let us look at an example. Having described reduction as being concerned with ontological concerns, Tim Crane states:

> What reduction needs, in addition, is the idea that the 'reduced phenomenon' is made more comprehensible or intelligible by being shown to be identical with the 'reducing phenomenon'. (Crane 2001, p. 54)

Unfortunately, when Crane briefly mentions how this idea of making things more comprehensible is to be spelled out, he merely repeats the Nagelian intuition that reduction is concerned with explaining one theory by another (ibid, p. 55).

So, why are 'water' and 'eau' cognates? Well, 'water' and 'eau' express the same concept and are cognates in this respect. Put differently: They are synonymous. The first observation can be stated as follows: If 'water' (derivatively) reduces to 'H_2O', then 'eau' reduces to 'H_2O'. (For the sake of clarity, I ignore questions concerning language-identity as a potential necessary condition for synonymy, which poses a mere technical problem.) We can generalize:

(Observation A)
If 'water' reduces to 'H_2O' then any expression being synonymous to 'water' reduces to any expression being synonymous to 'H_2O' (and to 'H_2O').

The second observation concerns the relation between expression-replacement of synonymous expressions and explanations of why these replacements count as reduction. Let us assume that we already know why 'water' reduces to 'H_2O', say, in virtue of *p*. Being able to give this explanation, we obviously have an appropriate explanation for why any expression being synonymous to 'water' reduces to any expression being synonymous to 'H_2O' and to 'H_2O', and why 'water' also reduces to any expression being synonymous to 'H_2O'; this is the case in virtue of *p*. So, the second observation is this:

(Observation B)
If 'water' reduces to 'H_2O' then any expression being synonymous to 'water' reduces to any expression being synonymous to 'H_2O' and to 'H_2O' for the same reason 'water' reduces to 'H_2O'.

Synonymy is defined on the level of conceptual content, neither on the level of properties or natural kinds – these being the referents of our sense-bearing expressions – nor on the level of syntax. Accordingly, 'water' and its cognates differ from 'H_2O' on the level of conceptual contents. Since ordinary intensions (functions from possible worlds to extensions) do not help specifying the relevant level – intensions are individuated too coarsely –, I shall use the term 'hyper-intensions' in order to refer to these conceptual contents.

Talking about hyper-intensions, one should be able to specify the identity-conditions of conceptual contents. Before I go on, I will define a notion of synonymy. In the last section, I shall briefly outline a more formal and more informative way that enables us to distinguish between different conceptual contents. For the moment, all we need is a pre-theoretic grasp of the notion of a conceptual content.

Synonymous expressions are *cognitively equivalent*. I follow Künne in putting the notion of Fregean *equipollence* (Frege (1969), p. 213) this way (every variable ranges over individuals):

(CognE)
Two sentences are cognitively equivalent iff, for any context x, nobody who fully understands them can take one of them to express a truth with respect to c without

> *immediately being ready to take the other to express a truth with respect to c as well.* (Künne 2003, p. 42 f.)

Künne argues correctly that (CognE) does not furnish us with a sufficient condition for sameness of senses of sentences. Every n-tuple of evidently true sentences is such that its members are cognitively equivalent without thereby expressing the same proposition. But what distinguishes the sense of '2+2 = 4' from the sense of 'Napoleon is a table or Napoleon is not a table'? Well, its constituents. But this will not help, because (i) '4^2 =16'and '2^4=16' have the same constituents (even though they are structured differently), they are cognitively equivalent and nevertheless express different thoughts, and (ii) talk about constituents requires an idea of how constituents are to be individuated.

Putting this problem aside for a moment, I will now outline an alternative idea – that the sameness of sense is mimicked on the level of sense-grasping physical entities. The physical state (-type) a person is in grasping a sense of a sentence is sufficient (even though presumably not necessary) for grasping that sense, given that determinism is correct and given a certain historical and nomological background. So, two sentences express the same sense iff a person, when grasping the content of one of these sentences, could have been (partly) in the same physical state when grasping the content of the other sentence. This tells us, in a somewhat idealized way, what it is for someone to grasp the same content in possibly different situations. Note that 'content' is ambiguous. According to one reading, it is lexical meaning that 'content' refers to. Context sensitive expressions have the same content without their tokens being necessarily synonymous. Therefore, we must additionally specify what kind of content we want to focus on. (CognE) excludes the possibility of classifying sentences having *merely* the same lexical meaning as being synonymous, because it introduces the notion of truth. Therefore, we should try to incorporate the notion of truth to get a proper definition of synonymy:

> *(SynDef)*
> *Sentence x is synonymous to sentence y iff*
> *there is a possible concept-grasping entity e having a physical basis, a state b, a time t, a deterministic world w and another deterministic world w* such that*
> *e is in b at t in w & e thereby believes that the content of x is true &*
> *e is in b at t in w & e thereby believes that the content of y is true &*
> *w and w* are law- and history-indistinguishable up to t (except features concerning the possibility of being related to different expressions at t).*

Here, we have fixed the notion of synonymy of sentences. Two non-sentential expressions are synonymous iff for every sentence forming operator working on these expressions, the resulting sentences are synonymous according to (SynDef).[4]

Now, let us come back to reduction. In order to answer the question why, say, *water* reduces to H_2O, we should have a closer look at the conceptual level, for it is at this level that the relevant relations between 'water' and, say, 'eau' on the one hand, and 'water' and 'H_2O' on the other are to be found. As (Observation B) points out: Since identity of the hyper-intensions of two terms suffices for reducing *for the same reason* (if these terms actually reduce to other terms) it is the level of conceptual contents that should be investigated to derive an appropriate definition of reduction. I think that this can be connected to the idea that reductions improve comprehensibility or intelligibility.

The idea of an alternative explanation of intertheoretical reduction is easy to sketch: Different conceptual contents present us with things differently, and sometimes, two conceptual contents present us with the same thing in different ways. The way the concept of water presents us with water is fundamentally different from the way the concept of H_2O presents us with water. This tells us in which respect two linguistic entities have to differ in order to reduce. But it does not yet tell us anything about the direction of the reduction-relation. So, this is my suggestion:

Some ways of presenting a thing might be more appropriate than others. The notion of appropriateness seems to require something with respect to which it can be truthfully attributed, that is: it needs a context. The basic idea is that, in a certain context, *the notion of H_2O is more appropriate than the notion of water*. This tentative characterization delivers a sound picture of reductionist claims, like: *(The surface property of being) water reduces to (the chemical property of being) H_2O* (an example of this kind can be found in Levine 1993, 131, f.), or *human pain reduces to c-fibre stimulation*. According to folk-chemistry, we can explain why mosquito larvae sink if we add washing-up liquid to the water they swim in: because the surface energy diminishes, and it was the surface energy that prevented the mosquito larvae from sinking. In chemistry we can explain events of the same kind

[4] I take this to follow from the notions of synonymy, concept-grasping physical entities and determinism.

talking about them as destruction of hydrogen bonds and about the occurrence of this destruction as being causally linked to the chemical structure of water and to the chemical structure of washing-up liquid. We change the way we talk about the same things, and science (partly) differs from ordinary language in that it delivers explanations explicitly or implicitly mentioning properties that are *causally relevant for the events or states of affairs the occurrence of which we try to explain*. For example, the property of having a hydrogen-molecule as a constituent is, so to speak, *implicitly mentioned* using the expression 'H_2O' in current English. This provides a hint for specifying what kinds of context are relevant: 'Water' reduces to 'H_2O' with respect to causal events, since the concept of H_2O presents us with water as having certain causally relevant properties, properties that are not picked out by the concept of water. Some concepts are *transparent* with respect to certain phenomena in that they pick out properties that are causally, or, more generally, explanatorily relevant with respect to these phenomena. Thus, H_2O is said to be effected by washing-up liquid in a certain way in virtue of having the property of being composed of oxygen and hydrogen. Accordingly, reduction seems to be a certain sort of explanation improvement. How can this idea be shaped more precisely?

We should take into account that reductionists claim that the reducing predicates (as parts of a theory) have the same intension the reduced predicates have, or, similarly, that the corresponding natural kind terms, like 'water' and 'H_2O' have the same intension. Generalizing, we can put the idea of reduction as explanation improvement as follows: If two concepts A (concept of H_2O) and B (concept of Water) as wholes pick out the same property (property of being water) and if A is more transparent than B with respect to an event-type E (formation/destruction of hydrogen bonds), then B is to be replaced by A with respect to E.

The notion of transparency can be defined defining the notion of *relevance of a property with respect to an event-type*:

> *(RelDef.)*
> *Property p is relevant for e iff*
> *events of type e take place at least sometimes in virtue of an instance of p.*

The relation in virtue of which the explanation relation holds can be causal in nature, but this is not necessarily the case. Consider a monist who

believes that the physical reduces to the mental. I think that we should take this to be a consistent belief – and even if it is inconsistent, it is not inconsistent in virtue of the notion of reduction employed here. But it would turn out to be inconsistent, if the notion of reduction were applicable only in cases where the criteria of appropriateness of a concept require causal relations between the properties and events or states of affairs that are to be explained. Clearly, our monist would not subscribe to the thesis that the causal relations out there are basic, but rather that some mental relations are basic. In a sense, the notion of reduction is topic-neutral.
Let me describe the relevant relation of *picking out* which holds between concepts and properties this way:

> *(ConcPropDef.)*
> *The sense x of expression y picks out property p iff there is a concept x* being a (proper or improper) constituent of x such that for every expression y* expressing x* the result of applying the operator 'the property of being (an) _' to y* refers to p.*

As for proper and improper constituents: The concept of hydrogen is a proper constituent of the concept of H_2O, whilst the concept of H_2O is an improper constituent of H_2O. Now, the notion of *being more transparent than* can be defined:

> *(TranspDef.)*
> *Concept y is more transparent than concept x with respect to event-type e iff*
> *y consists of concepts picking out more properties that are relevant for e than y does.*

And this makes for a definition of reduction, taking into account that the reducing and the reduced concept as wholes (their improper constituents) have to pick out the same property, or that the corresponding expressions have the same intension, just like the concept of water and the concept of H_2O as wholes presumably pick out the same property and 'the property of being water' and 'the property of being H_2O' have the same intension:

> *(RedDef)*
> *Expression x reduces to expression y with respect to event type e iff*
> *(i) the conceptual content of y is more transparent with respect to e than the conceptual content of x and*
> *(ii) x and y have the same intension.*

In this sense, reduction is a sort of explanation-improvement. It is expression replacement to the effect that the conceptual contents of expressions become more transparent with respect to the explanatorily relevant properties. [5]

Explanation comes in degrees. Some explanations are better than others. An explanation of hydrogen-bond-involving events involving the concept of H_2O is better than an explanation just using the concept of water. Nevertheless, an explanation of this latter type can be true, or correct. We now can answer the question in which sense reduction is concerned with explanation of one theory by another:

An explanation of an event of destruction of hydrogen-bonds in terms of hydrogen-bonds and H_2O can be used to explain why the corresponding explanation expressed in terms of folk-chemistry is true, if we add information of the following kind: *Water is nothing but H_2O*.

One might be a little bit upset by the fact that the reduction-relation turns out to be a four-place-relation (as far as expressions are concerned) or a

---

[5] Similar considerations have inspired some philosophers to investigate levels of explanations as opposed to ontological levels. Unfortunately, these philosophers tend to conflate both ways of individuating levels of description. Peter Smith, for example, argues that the different levels of explanation corresponding to, say, psychology and neuroscience differ with respect to the taxonomies employed by these theories (Smith 1992). These taxonomies seem to be classifications on the level of event-types or properties, not on the level of conceptual contents. Adrian Cussins, who does not aim at illuminating the notion of reduction, but rather presupposes such a notion and tries to establish a notion of non-reductive *naturalism*, claims that we should distinguish scientific levels according to understanding conditions (Cussins 1992, p.182) – which comes close to a level of conceptual contents – but he himself, at least sometimes, describes the difference of levels in terms of properties (ibid. p.179). Robert van Gulick seems to follow a similar line of thinking: He explicitly argues that the difference between a theory of the mental and neuroscience partly consists in the fact that the former contains *concepts* that cannot be *expressed* within the framework of the latter (van Gulick 1992, p. 163). In a more recent paper he describes reduction as the task of gaining a 'mode of access with the alleged reducing representations' that can be gained with the reduced representations (van Gulick 2001, p.14). Different *modes of access* correspond to different pragmatic dimensions of user-theory-relations (*cf.* van Gulick 1992, p. 166 f). Even though van Gulick discusses questions of theory-individuation in terms of concepts, the underlying metaphysics makes it difficult to understand the role these conceptual contents and modes of access play. He argues that even though the mental belongs to the domain of the physical, mental properties are not to be identified with physical properties. What the conceptual differences between both theories add to this metaphysical difference is hard to uncover. Since different theories can, according to van Gulick, deliver the same mode of access, a mode of access cannot deliver a criterion for concept-individuation and, therefore, not illuminate the role the conceptual differences play; different theories should at least contain different conceptual contents. Otherwise, the notion of reduction could easily be defined in terms of conceptual-content preserving translation.

three-place-relation (as far as conceptual contents are concerned), that is, that some term or concept reduces to another *with respect to certain states of affairs or events*. Isn't there reduction *simpliciter*? Let me briefly argue against the possibility of reduction *simpliciter*. The notion of reduction seems to comprise an element of appropriateness of a concept. The concept of H_2O is more appropriate than the concept of water. But nothing is appropriate *simpliciter*. Even though we can perfectly judge questions to be inappropriate, period – that is, we can use the one-place predicate `_is (in)appropriate' to generate a well-formed sentence – the property of being (in)appropriate is nevertheless a relational property. Something is appropriate or inappropriate with respect to something. And just like no one is in love *simpliciter*, no expression reduces to another *simpliciter*. Nevertheless, there is an interesting kind of reduction that comes close to reduction *simpliciter*, a kind of reduction we should call more appropriately `universal reduction': A reduces to B *universally* iff for any event or state of affairs that possibly occurs in virtue of the property/the substance/the kind which both `A' and `B' refer to or can be used to ascribe, A reduces to B. This definition makes universal reduction conceptually dependent on reduction as outlined above.
Before I go on, let me briefly comment on the question of how this sort of reduction relates to the metaphysical/explanatory-relation. I think that this sort of reduction is a sort of explanation-improvement in virtue of its being concerned with metaphysics. It is concerned with metaphysics not only in that it states identity claims. Furthermore it makes the criteria for explanation-improvement dependent on the *nature* of things – this nature corresponds to the different modes under which things are presented to us by different concepts. In causal contexts, `water' reduces to `H_2O' because the concept of H_2O informs us better about the causal nature of water, unlike the concept of water.

Can we connect these insights to inter-theoretic relations? In order to do so, let me briefly comment on cases like the phlogiston-oxygen case. The relevant relation between the phlogiston theory and the oxygen theory is in many respects a mere historical one. Nevertheless, many philosophers would be ready to accept that, in one sense or another, the phlogiston theory actually has an explanatory power being closely related to the explanatory power of the oxygen theory (or that it fits a lot of data the

oxygen theory fits). I take this to mean that, if the phlogiston-theory were true, then it would explain at least part of what the oxygen-theory explains. But in fact it did not explain anything, for nothing can be explained by referring to a false theory. Mild reflection on the use of 'because' shows this: We are not ready to acknowledge that a sentence of the form 'Q because P' expresses a truth if P is wrong, and we do so for conceptual reasons. So, if the phlogiston theory did not explain anything, the interesting relation between the oxygen and the phlogiston theory seems to be based on counterfactual truths concerning the realm of phenomena both theories *would* explain. But for every kind of phenomena (e.g. those that are explained by the oxygen theory), there are infinitely many theories that *would* explain these phenomena (equally saving the data, though perhaps not equally well). It sounds odd to say that all these theories *reduce* to the oxygen theory in the sense of a theory-shift, a notion that guided many philosophers of science through their investigations of the notion of intra-level reduction; therefore, if the phlogiston theory reduces to the oxygen theory, then this is to a relevant degree due to (important) historical and psychological reasons. A similar story can be told about corrections in the base theory; if a reduced theory is corrected by or in the base theory, it contains false parts the base theory does not contain. These parts have the same status as the phlogiston theory: They explain nothing and are replaced in the base theory. I mention this notion of theory-succession-reduction because this concept is sometimes conflated with the sort of reduction I try to outline.[6]

To define a non-historical notion of intertheoretic reduction, we should accept that false (parts of) theories do not reduce to true (parts of) theories. To sketch the basic idea, consider the following (schemas of) general claims which I use as dummies for (schemas of) sentences expressing natural laws whatsoever:

1) $\Box_{nom} \forall x$ (x is water $\rightarrow$ Gx)

[6] Here are two examples: David Charles and Kathleen Lennon describe reduction as follows: 'Reductionist accounts aim to show that where *we thought* we had two sets of concepts, entities, laws, explanations, or properties, we in fact have only one […]' (Charles & Lennon 1992, p.2, my italics). Tim Crane also refers to epistemic subjects: '[W]e start off with the 'target' entity, X, and find a reason for identifying X with Y.' (Crane 2001, p. 54) This version of reduction seems to underlie slogans like 'if As reduce to Bs, then As do not exist over and above Bs', since this slogan is telling just in case *someone actually thought* that As exist over and above Bs.

2) $\Box_{nom} \forall x$ (x is $H_2O \rightarrow G^*x$)

Assume that both sentences express truths. The former reduces to the latter, because there is a function mapping the non-logical constants of the former onto non-logical constants of the latter. This function is governed by the following principle:

> *(TheorRedPrinc)*
> *Map all the relevant sense bearing constants of the reducing theory onto the constants they reduce to – in the abovementioned sense – or to themselves in the base theory.*

A theory reduces to another iff such a function can be found. Now, we can define inter-level theory-reduction as follows:

> *Theory A reduces to theory B iff A has a logical structure of B* (B* being identical to B or to a substructure of B) and there is a set of functions such that these functions map all the constants of A onto themselves in B* or, if they do not occur in B*, onto constants they reduce to.*

Whether or not we want to allow extensions of A in B (such that B contains expressions being neither contained by A nor being the reducers of any member of A), as this definition does, is rather a matter of taste or purpose than of conceptual considerations concerning reduction. Note that the notion of a theory used here must not be identified with classical views of theories. The theories I talk about are interpreted theories in the sense that their sense-bearing constituents have a hyper-intension. (For a sketch of a model of hyper-intension, see below.)
This sort of theory reduction can be applied not only to reduction of folk-theory to scientific-theory, but also to intra-scientific inter-level-reductions. Biology, for example, reduces to chemistry in this sense iff (i) the ontology of chemistry contains a subclass being equivalent to the ontology of biology and (ii) the conceptual content of the language of chemistry presents us with this ontology in a more appropriate way – that is: it presents us with more causally relevant properties.[7]

[7] Let me add another observation: If dependence-relations between properties are relevant for the notion of reduction, then it is with respect to the properties corresponding to the mode of presentation, not with respect to the properties the concepts pick out as wholes. It is not the case that the property of water depends (in an interesting sense of 'depends') on the property of H_2O. But the properties corresponding to the mode of presentation of the concept of water, maybe some phenomenal properties, are dependent on the properties

3. Semantic approaches

According to classical semantic approaches to intertheoretic reduction, reduction is not based on syntactic properties. Nevertheless, these approaches face another problem. As Suppe pointed out, we should regard 'theory individuation issues as make or break for any account of theories' (Suppe 2000, p. 109). According to my view, theories (or better: the entities in virtue of which theories reduce) should be individuated very fine grained; it is within the realm of Fregean senses that we have to look for the individuating criteria. If we do not, a water-theory does not semantically differ from an H_2O-theory, and, therefore, cannot be reduced to the latter. As long as conceptual differences, that is: differences on the level of hyper-intensional entities, cannot be explained in terms of model-theory, model-theoretic approaches fail to explain what theory reduction consists in. This is because of the fact that model-theory takes the primary role of predicates, like '_ is water' and '_ is H_2O' to consist in picking out Carnap-inspired intensions rather than hyper-intensions. They are unable to mark the difference between, say, the proposition that is expressed by 'iron is a metal' and the proposition that is expressed by 'Fe is a metal'. Nevertheless, one might hope, as I do, that there is a way of developing individuating criteria for conceptual contents using set-theoretical notions. There are at least two accounts on offer, both of which lead to untenable results. Since an appropriate discussion of these accounts would take too much space, I will just point to the main problem they face in a footnote and briefly sketch an alternative.

This alternative starts with a pre-theoretic understanding of the notion of a mode of presentation and rests on the assumption that a concept's mode of presentation is what makes for the concept's identity. I take a mode of presentation to be something that presents us with the world as if there are objects/is an object, and these objects/this object as having or lacking certain properties. Furthermore, for the sake of simplicity, let us take properties to be functions from worlds to extensions (the approach will work for any model that describes properties in terms of set-theory).

---

corresponding to the mode of presentation of the concept of H_2O. Water is tasteless for us (at least in part) *in virtue of* having its chemical structure.

Now, if we take the function of conceptual contents of a singular term or a predicate to have a certain mode of presentation, and if we subscribe to the thesis that properties are individuated by functions from possible worlds to extensions, and if we accept that for any concept A, for any concept B, B=A iff A and B have the same mode of presentation, then for any basic concept, for any expression expressing this concept, the expression's intension determines the basic concept that is expressed. To put it differently: No two expressions that have the same intension and express basic concepts can express different concepts. Basic concepts, being unanalyzable, present us with things as if there is/are something/some things and this thing/these things as having exactly one property. I take this to be a conceptual point about the notion of a mode of presentation and the notion of a basic concept.

Now, we can take the conceptual content of an expression that expresses a complex conceptual content (its hyper-intension) to be a structure of intensions corresponding to basic concepts forming the complex concept. This structure contains some of the intensions determining (or being equal to) the properties that are *picked out* by the complex concept. These concepts have to be structured appropriately: If C is a concept being composed of the three basic concepts *f*, *f**, and *f*+ such that something falls under C in a world w iff (i) it falls under *f* and *f** in w or (ii) if it falls under *f*+ in w, then the hyper-intension of an expression expressing C is the following structure:

<<*f*, *f**, *f*+>, {x, y, y*, y+, w: (x falls under y in world w & x falls under y* in world w) ∨ x falls under y+ in world w}>

An entity x *falls under* a function from worlds to extensions *f* in a world w iff *f* assigns value v (an extension) to w and x is a member of v, that is: if it has the corresponding property in w. This hyper-intension determines the intension and the extension of any expression having this hyper-intension.[8]

[8] The abovementioned two alternatives do not start with basic *concepts*, but with basic or simple *expressions*. David Chalmers' version of two-dimensionalism allows for a model of something very similar to hyper-intensions. A sentence's hyper-intension could be modelled as a structure of primary intensions (functions from scenarios to extensions) of the sentence's constituents. (Chalmers 2006). The problem is that a sentence's constituents (the simple expressions) not necessarily mimic the semantic structure they have: A simple expression can express complex conceptual contents. Chalmers introduces these structured intensions in order to distinguish between the semantics of pairs of expressions of the

So, a hyper-intension of an expression E is a structure containing functions from possible worlds to extensions (corresponding to the properties the primitive concepts pick out) and a relation that defines how the basic concepts relate to the members of the concept's extensions in worlds.
Basic sentences can be taken to contain two such structures – the structure of a singular term and of a predicate:

<<singular term>,<predicate>>

Complex sentences are n-tupels of pairs of hyper-intensions of singular terms and predicates and a relation that defines the composition of these basic sentences' hyper-intensions. I do not have the space here to go into this in detail. But I hope that this roughly sketched drawing illuminates the relation between this sort of semantics and the notion of reduction outlined above. If we take theories to be syntactic entities that have hyper-intensions, we can describe reduction as follows: A reduced theory's hyper-intensions deliver new functions or properties, and these functions or properties are more relevant than the properties delivered by the hyper-intensions of the reducing theory, even though the theories' intensions are equivalent. According to the paradigmatic cases that guided the discussion, the mind *reduces* to the physical in this sense, if it reduces at all.

Conclusion

I briefly outlined a serious problem which many accounts of intertheoretic reduction face: They cannot *explain why* a theory reduces to another. The

---

following kind: '7+3' and '10'. Let me stipulate that '10*' is a simple expression that has the same meaning as '7+3'. The difference between the semantic value of '10*' and the semantic value of '10' cannot be drawn along the line of structured primary intensions of simple expressions – the simple expressions have the same primary intension. Furthermore, both expressions have the same extensions such that an alternative way Chalmers suggested – taking senses of basic expressions to be pairs of primary intensions and extensions (ibid.) – is blocked.

Cresswell's account of structured propositions has several problems stemming from the same source (Cresswell 1985). According to Cresswell, a proposition (the second relatum of, say, a belief-relation) can be modelled as a structured set of ordinary intensions of simple expressions. Therefore, we are unable to mark a difference not only between the semantic values of '10*' and '10'. Additionally, the conceptual contents of 'iron' and 'Fe' turn out to be identical, since both expressions are (i) simple and (ii) have the same intension. But they have different conceptual contents. So, we should not focus on simple expressions and their senses to look for the furniture of structured meanings.

account outlined in the present paper can answer this question: The concepts expressed by elements of the base theory are more relevant with respect to the phenomena the theory seeks to explain than the elements of the reduced theory. Furthermore, this account can explain the epistemic relevance of this kind of reduction: It is the project of making explanations more appropriate by pointing to the explanatorily relevant properties which were not available in the reduced theory.

Note that the concept of reduction does not include a certain stance on the question of what actually is the fundamental level. An idealist can use the term in the same way to state his theses as the physicalist can do. Therefore, no physical concept of causation enters into this concept, nor does the concept presuppose a certain hierarchy of scientific levels.

References

- Bickle, John (1998). *Psychoneural Reduction: The New Wave*. (Cambridge, MA: MIT Press)
- Chalmers, David (2006). Two-Dimensional semantics. (In: Lepore, E. & Smith, B. (Eds.) *Oxford Handbook of the Philosophy of Language*. Oxford: Oxford University Press.)
- Charles, David & Lennon, Kathleen (1992). Reduction, Explanation and Realism (Introduction). (In: Charles, D. & Lennon, K. (Eds.) *Reduction, Explanation and Realism* (pp. 1-19). Oxford: Clarendon Press)
- Cresswell, Max J. (1985) *Structured Meanings: The Semantics of Propositional Attitudes*. (Cambridge, MA: MIT Press)
- Fodor, Jerry A. (1974). Special Sciences. *Synthese, 28,* 77-11.
- Fodor, Jerry A. (1981): *Representations. Philosophical Essays on the Foundations of Cognitive Science*. (Cambridge, MA: MIT Press.)
- Frege, Gottlob (1969). *Nachgelassene Schriften.* (Hamburg: Meiner.)
- Hooker, Clifford A. (1981): Towards a General Theory of Reduction. Part I: Historical and Scientific Setting. Part II: Identity in Reduction. Part III: Cross-Categorial Reduction. *Dialogue, 20*: 38-59, 201-236, 496-529.
- Kim, Jaegwon (1993). Multiple Realization and the Metaphysics of Reduction. (In: *Supervenience and Mind, selected philosophical Essays* (pp. 309-335). Cambridge: Cambridge University Press.)
- Künne, Wolfgang (2003). *Conceptions of Truth.* (xford: Oxford University Press.)
- Levine, Joseph (1993): On Leaving Out What It's Like. (In: Davies, M. & Humphreys, G. (Eds.). *Consciousness: Psychological and Philosophical Essays* (pp.121-136). Oxford: Blackwell)
- Marras, Ausonio (2992). Kim on Reduction, *Erkenntnis,* 57, 231-257.

- Marras, Ausonio (2006). Emergence and Reduction: Reply to Kim, *Synthese 151*, 561-569.
- Moulines, C. U. (1984): Ontological Reduction in the natural Sciences. (In: Balzer, W. Pearce, D. A. and Schmidt, H. J. (Eds.), *Reduction in Science* (pp. 51-70). Dordrecht: Reidel.)
- Nagel, E., 1961, *The Structure of Science.* (London: Routledge and Kegan Paul)
- Putnam, Hilary (1975): *Philosophical Papers. Vol. 2: Mind, Language and Reality*. (Cambridge: Cambridge University Press.)
- Schaffner, Kenneth (1993). *Discovery and explanation in biology and medicine*. (Chicago: University of Chicago Press.)
- Peter Smith (1992). Modest Reductions and the Unity of Science. (In: Charles, D. and Lennon, K. (Eds.). *Reduction, Explanation and Realism* (pp. 19-45). Oxford: Clarendon Press)
- Suppe, Frederick (2000): Understanding Scientific Theories: An Assessment of Developments, 1969-1998. *Philosophy of Science,* 67, 102-115.
- Van Gulick, Robert (1992). Nonreductive Materialism and the Nature of Intertheoretical Constraint. (In: Beckermann, A., Flohr, H. and Kim, J. (Eds.). *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism* (pp.157-180). Berlin/New-York: de Gruyter)
- Van Gulick, R. (2001). Reduction, Emergence and Other Recent Options on the Mind/Body Problem: A Philosophic Overview, *Journal of Consciousness Studies*, *8*, 9/10, 1-34.