The Reductive Explanation of Boiling Water in Levine's Explanatory Gap Argument

Max Seeger

In this paper I examine a paradigm case of allegedly successful reductive explanation, viz. the explanation of the fact that water boils at 100°C based on facts about H_2O. The case figures prominently in Joseph Levine's explanatory gap argument against physicalism. I will study the way the argument evolved in the writings of Levine, focusing especially on the question how the reductive explanation of boiling water figures in the argument. It will turn out that there are two versions of the explanatory gap argument to be found in Levine's writings. The earlier version relies heavily on conceptual analysis and construes reductive explanation as a process of deduction. The later version makes do without conceptual analysis and understands reductive explanations as based on theoretic reductions that are justified by explanatory power. Along the way I will stress the crucial role of the bridge principles which are neglected in the explanatory gap literature.

Introduction

Within the discussion of the mind-body problem one of the major arguments against physicalism has been the explanatory gap argument. Roughly, it states that there can be no reductive explanation of phenomenal consciousness based on physics. The argument comes in different flavors:[1] While David Chalmers draws the metaphysical conclusion that physicalism is false, Thomas Nagel argues for the somewhat conceptual conclusion that, if physicalism is true, we cannot coherently conceive of it. Joseph Levine, who coined the term "explanatory gap", argues for the epistemological conclusion that we cannot prove physicalism.

The most basic thesis that proponents of the explanatory gap argument will have to defend is the thesis of the epistemic gap, which says: while we can in principle reductively explain physical macro-properties (others speak of superficial or surface properties) based on microphysics, we cannot in the same sense reductively explain states of phenomenal consciousness based on physics.

In his 1993 Levine introduces a paradigm example to illustrate the alleged disanalogy: He sketches the reductive explanation of the disposition of water to boil at 100°C at sea level in terms of chemical facts about H_2O and general laws of chemistry.[2] He goes on to argue that it is not possible to explain phenomenal consciousness in the same way as it is possible to explain the boiling point of water. It is this paradigm example, the reductive explanation of boiling water in terms of H_2O, that I am most concerned with in this paper. More precisely, I will try to figure out how the alleged reductive explanation exactly is supposed to work and how it figures in the argument.

[1] Recent formulations have been put forward most prominently in Nagel 1974, Kripke 1980, Levine 1983 and 1993, Chalmers 1996, Jackson 1998 and Chalmers and Jackson 2001.

[2] It is an unfortunate artefact of this example that the explanandum arguably is true by definition: Degrees Celsius are (or, at any rate, were originally) defined in terms of boiling and freezing points of water. Thus it can be argued that the explanandum is an a priori truth (cf. Kripke 1980, 56; Kripke uses this as an example of an a priori contingent truth.). We can disregard this fact about the Celsius-unit and focus on the question at issue: Why is it that water boils at exactly the temperature it does? Also, Levine actually states his argument using the Fahrenheit unit, for which the problem does not arise.

I follow Levine's writings chronologically, thereby describing the development of his argument. It turns out that two versions of the explanatory gap argument have to be distinguished: According to one version (conceptual argument) reductive explanations require an a priori entailment from explanans to explanandum. This entailment is allegedly facilitated by a priori conceptual analysis. According to the other version (empirical argument) reductive explanations are based on empirical identity statements. I will propose to read Levine's first expression of the argument in his 1983 as a conceptual version. The *locus classicus* of the argument is Levine's 1993, which again is usually interpreted to express the conceptual argument. His later writings (1998, 1999, 2001) clearly express the empirical argument. It is somewhat perplexing that Levine never made his change of position explicit. In his 1999 he actually contrasts the two versions of the argument, saying that he professes the empirical argument, not the conceptual, but withholding that he used to profess the latter. One way to make sense out of this is to question whether he ever really did profess the conceptual argument at all. The controversial thesis of this paper will be just this: I will challenge the standard interpretation of Levine 1993 in arguing that it can be read to contain both versions. Putting it mildly, Levine's 1993 is indecisive as to which version of the argument is expressed, putting it harshly, it is inconsistent.

1983: The Explanatory Gap Argument

Levine first expressed the explanatory gap argument in his "Materialism and Qualia: The Explanatory Gap" (1983). The main thesis of the paper is that the explanatory gap is of an epistemic nature and does not license metaphysical conclusions. Levine's purpose, as he puts it, is to transform Kripke's anti-materialist argument from a metaphysical into an epistemological one. However, I will not concern myself with the dispute between Levine and Kripke, but just look at Levine's positive argument for the explanatory gap. Here is how it goes.

The thesis of physicalism is—depending on one's particular version—that every phenomenon is either identical with or reducible to physical phenomena. So any mental state, for example pain, is identical with or reducible to a physical state, for example the firing of C-fibers. But even if we assume the truth of the identity-statement 'pain = C-fiber firing (Cff)', it wouldn't really be satisfactory for the following reason: There seems to be an intuitive difference between identity-statements involving phenomenal states, such as 'pain=Cff', and regular scientific identity-statements such as 'heat = molecular kinetic energy (mke)'. While 'heat=mke' is "*fully explanatory*", 'pain=Cff' is not, it leaves an "*explanatory gap*" (Levine 1983, 357). This observation, that there are two different kinds of identity-statements, is the starting point for the explanatory gap argument.

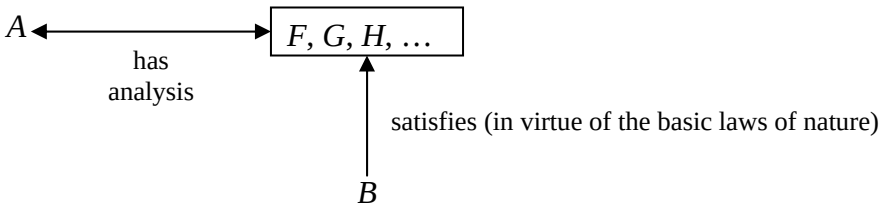
What exactly does it mean for an identity-statement to be fully explanatory? 'Heat=mke' is fully explanatory in the intuitive sense that "whatever there is to explain about heat is explained by its being the motion of molecules" (1983, 359). While Cff can in the same sense explain pain-*behavior*, i.e. the physical causes and effects related to pain, it seems intuitively clear that it cannot explain the phenomenal aspect of pain: "what is left unexplained by the discovery of C-fiber firing is *why pain should feel the way it does*!" (1983, 357; Levine's emphasis)

The explanatory gap argument is mostly concerned with capturing this intuitive difference in a theoretic framework. Levine puts it this way: 'heat=mke' is fully explanatory because heat *can be reduced to* mke, or in other words, mke *reductively explains* heat. On the contrary, pain does not reduce to Cff, or, Cff does not reductively explain pain, which means that the identity statement 'pain=Cff' is not fully explanatory. Now, what does it mean for mke to reductively explain heat? Levine's notion of reductive explanation is best captured in a two-stage account. First, we analyze the concept of the phenomenon to be reduced (explanandum) by citing

necessary and sufficient conditions under which the concept applies. Second, we show that the reducing phenomenon (explanans) fulfills these very conditions. We then see, why, given the explanans, the explanandum *must* be the case. In other words: Given the explanans, it is inconceivable that the explanandum not be the case.

Levine illustrates his account of reductive explanation using the reduction of heat to mke as an example. The concept of heat is analyzed by Levine by stating the causal role of heat. Heat is the "phenomenon we experience through the sensation of warmth and cold, which is responsible for the expansion and contraction of mercury in thermometers, which causes some gases to rise and others to sink, etc." (1983, 355) And this causal role, we know by empirical discovery, is filled by the motion of molecules. Once we understand, for example, what it means for mercury molecules in thermometers to have a certain kinetic energy, it becomes clear that it is this energy which is responsible for the expansion or contraction of the mercury.

The two stage process can be illustrated by the following diagram.



The explanandum *A* is analyzed as that phenomenon, to which the predicates *F*, *G*, *H*, etc. apply. The explanatory base *B* satisfies those predicates in virtue of the basic laws of nature.

It must be conceded that the structure of reductive explanation sketched above is not *explicitly* described or endorsed by Levine in his 1983 (even though it is later, in his 1993). However, several passages imply that this is what he must have in mind. Consider the following.

> What is explanatory about ['heat=mke']? […] The explanatory force of this statement is captured in statements like (2') above [(2') says that mke "is responsible for the expansion and contraction of mercury in thermometers … causes some gases to rise and others to sink, etc." see above]. (2') tells us by what mechanism the causal functions we associate with heat are effected. It is explanatory in the sense that our knowledge of chemistry and physics makes intelligible how it is that something like the motion of molecules could play the causal role we associate with heat. Furthermore, antecedent to our discovery of the essential nature of heat, its causal role, captured in statements like (2'), exhausts our notion of it. Once we understand how this causal role is carried out there is nothing more we need to understand. (1983, 357)

In this description we can identify two steps. First, conceptual analysis tells us that heat is what fills the causal role of heat, and what that causal role is. Second, empirical description shows that mke fills just that role.

Concerning the conceptual analysis, it is of outmost importance that our notion of heat is "exhausted" by the causal role of heat. This is not so for concepts of mental states: "there is more to our concept of pain than its causal role, there is its qualitative character, how it feels" (1983, 357). So our analysis of 'pain' yields that pain fills a certain causal role and, in addition, has a certain qualitative character: it feels painful. But this by itself doesn't preclude pain from being reductively explainable. If we could show how it were that Cff fills that causal role and also how it were that Cff has a certain qualitative character, we'd have a reductive explanation of pain in terms of Cff. But the later, showing that Cff (or any other neurological state) has a certain qualitative character, seems to be impossible.

Concerning the empirical description, we may ask: how do we show that mke fills the causal role of heat? Levine says that "our knowledge of chemistry and physics makes it *intelligible*" (ibid., my emphasis) how mke plays this very role.

> What we need is an account of what it is for a phenomenon to be made *intelligible*, along with rules which determine when the demand for further intelligibility is inappropriate. (1983, 358; Levine's emphasis)

This means, once we have described the situation in chemical or physical terms, it is no longer intelligible that the temperature be different than it actually is. Generally, explaining something means making something intelligible. Another way of putting the idea is this. Once we understand the laws of chemistry and physics, when we know that some substance has a certain molecular kinetic energy, it is inconceivable that this substance should have a different temperature than it actually does.

Again, this can be contrasted to the case of pain. While a physical description of someone in pain can make it intelligible that she *behaves* the way she does, the qualitative aspects cannot be captured in physical terms. We can easily imagine someone's C-fibers firing without her being in pain.

> If there is nothing we can determine about C-fiber firing that explains why having one's C-fiber's fire has the qualitative character that it does […] it immediately becomes imaginable that there be C-fiber firings without the feeling of pain, and *vice versa*. (1983, 359)

The 'vice versa' should be taken to refer to the "if p then q"-structure of the statement, thus licensing the biconditional: Cff explains pain iff it is not imaginable that there is Cff but no pain.

The most obvious objection against the explanatory gap argument concerns the practice of scientific explanation. Levine relies heavily on a notion of reductive explanation that resembles the DN-model of scientific explanation. Opponents of the argument may object that deduction does not figure in the practice of scientific explanation in the way it is presupposed by Levine's notion of reductive explanation. Remember that the explanatory gap argument proceeds by analogy: mental states cannot be reduced to physical states in the same way physical macro-properties can be reduced to physical micro-properties. Opponents say that in scientific practice physical macro-properties are not in fact being reduced to micro-properties in the way sketched by Levine. However, this objection fails to seriously harm the argument for the following reason. Even if the reductive explanations sketched by Levine do not resemble actual explanations in scientific practice the argument may stand. All that Levine needs is that the reductive explanations can be given in principle. Rather than claiming that his account of reductive explanation resembles scientific practice, Levine assumes that the thesis of Materialism implies the explainability of all objects and properties that are not themselves basic.

> Materialism, as I understand it, implies explanatory reductionism of at least this minimal sort: that for every phenomenon not describable in terms of the fundamental physical magnitudes (whatever they turn out to be), there is a mechanism that is describable in terms of the fundamental physical magnitudes such that occurrences of the former are intelligible in terms of occurrences of the latter. (1983, 358f.)

The fact that they cannot be given for mental states raises a problem for physicalism. However, against this reply it may also be objected that physicalism need not imply this kind of explanatory reductionism. That is, the ontological claim that everything is physical does not imply the epistemological claim that we can explain everything there is in physically basic terms.

A second objection concerns brute facts. Could it be that psycho-physical identities are basic identities that need no further explanation, in the sense of brute facts? Again, Levine offers an

illustrative example: the laws of gravity explain falling bodies, but the laws of gravity themselves need no further explanation. Is it a viable strategy for the materialist to claim that psycho-physical identities are basic in a similar sense? Levine clearly negates this option. The main reason seems to be that brute facts should be expected on micro-levels, but not on macroscopic levels. Materialism, as Levine puts it, implies "that brute facts […] will not arise in the domain of theories like psychology." (1983, 359) The same point has been expressed above by saying that the explanatory basis *B* satisfies the analysis in question *in virtue of the basic laws of nature*. If we simply postulated 'pain=Cff' as a law of nature, Cff would indeed satisfy the analysis of pain in virtue of a law of nature. However, this would not be deemed a basic law and the move would seem ad hoc and not very explanatory. We do not want to postulate laws at the macro-level for the reasons just rehearsed. We are looking for general and basic laws of nature.

1993: The Reductive Explanation of Boiling Water

In his 1993 "On Leaving Out What It's Like" Levine elaborates the argument from 1983, especially the theoretical framework regarding his notion of reductive explanation. The one major difference, which is most relevant to my paper, is that he changed the paradigm example of successful reductive explanation; instead of 'heat=mke' he now discusses 'water=H_2O'. This move, as I will argue below, made things a lot more complicated. In this section I will present *On Leaving Out* in its standard interpretation, namely as a version of the conceptual argument. To do this, I briefly present Levine's somewhat more articulate theory of reductive explanation which I then illustrate in greater detail using the example of the boiling point of water. In the last section I will challenge this interpretation and argue that *On Leaving Out* already contains the core ideas of the empirical argument.

Reductive Explanation

The general structure of the argument remains unchanged: Physical macro-properties can be reductively explained in terms of physical micro-phenomena, whereas mental phenomena cannot. This difference presents a challenge to physicalism. For the challenge to be more than intuitive, one has to spell out what it means for some phenomenon *A* to be reductively explained by some phenomenon *B*. Levine now explicitly declares reductive explanation to be a two-stage process:

> Stage 1 involves the (relatively? quasi?) [sic] a priori process of working the concept of the property to be reduced 'into shape' for reduction by identifying the causal role for which we are seeking the underlying mechanisms. Stage 2 involves the empirical work of discovering just what those underlying mechanisms are. (1993, 132)

Taken at face value, several points are noteworthy. First, the required conceptual analysis is a more or less a priori process. Whether the required analyses are really to be had a priori is a controversial matter. Levine seems to have his own doubts, as he speaks of it as a "relatively" or "quasi" a priori process. Second, it is properties that are reductively explained, not substances, as one may be lead to think by Levine's concern for the identity statement 'water=H_2O'. And third, the conceptual analysis is given in terms of a causal role. We already saw that this is the case in standard examples such as heat/mke, but it is an open question whether this should be generalized.

But should we take this quote literally? Does reductive explanation require that the concept in question is analyzable in terms of a causal role? At another point he speaks of the "superficial

properties by which we identify water." (1993, 131) This suggests that other properties than the causally relevant ones do figure in the analysis as well. Also, if only those phenomena could be reductively explained, that are analyzable in terms of causal role, then all Levine would have to show is that mental states cannot be analyzed exhaustively in terms of causal role. This, however, can be considered a commonplace in the debate. (Of course, there are still some functionalists who claim that mental states can be understood in terms of causal roles.) Levine, on the contrary, tries hard to show that the phenomenal quality of mental states cannot be made intelligible by physical states. This shows that the exhaustive analysis in terms of causal role is not required for reductive explanation. [3]

Levine speaks of the underlying mechanism that fills a causal role. But how do we know that some mechanism actually fits some particular role? In other words, what does it take for the empirical description to be an *explanation*? This question concerns the epistemic relationship between explanandum and explanans.

> [A] reduction should explain what is reduced, and the way we tell whether this has been accomplished is to see whether the phenomenon to be reduced is *epistemologically necessitated* by the reducing phenomenon, i.e. whether we can see why, given the facts cited in the reduction, things must be the way they seem on the surface. (1993, 129; my emphasis)

So the notion of explanation is being spelled out in terms of epistemological necessitation. In Levine's account this means, that given the truth of the explanans it is *inconceivable* that the explanandum be false. That is, *B* epistemologically necessitates *A* iff it is inconceivable that *B* and not-*A*. I think we can safely say that this amounts to the following claim: *B* explains *A* iff the material conditional $B \rightarrow A$ is known to be true on solely logical and conceptual grounds.[4] This is the claim that there is an a priori entailment from *explanans* to *explanandum*.

Thus, to show that our chemical theory of H_2O reductively explains the boiling point of water Levine needs to prove that the conditional "If our chemical theory of H_2O is true, then water boils at 100°C" can be known to be true based solely on logical and conceptual grounds. To put it in his own words, he has to show that it is inconceivable that our theory of H_2O is true, and yet water doesn't boil at 100°C. However, he doesn't seem to claim that this can be shown. What he does say is this: "it is not conceivable, I contend, that H_2O should fail to manifest these properties [i.e. the macro properties of water]" (1993, 128; my emphasis) or "it is inconceivable that H_2O should not boil at 212°F at sea level" (1993, 129; my emphasis). One may thus wonder: what exactly is it that can be reductively explained by the theory of H_2O – the boiling point of *water*, or the boiling point of H_2O?

Of course, strictly speaking the boiling point of water *is* the boiling point of H_2O, so literally this question is nonsensical. One and the same fact may be represented in different ways, but what gets explained are facts, not representations. Still, since explanation is an epistemological

[3] On the other hand, Michael Pauen in an introduction to the explanatory gap argument argues that only those phenomena can be reductively explained, the concepts of which are analyzable in terms of a causal role: "Die entscheidende Voraussetzung für diesen Übergang [zwischen alltagssprachlicher und micro-physikalischer Ebene] besteht darin, daß wir das, was wir mit den alltagssprachlichen Ausdrücken ‚Wasser' oder ‚Eis' meinen, in abstrakte Beschreibungen übersetzen können, in denen nur noch von einer Rolle in bestimmten Ursache-Wirkungsbeziehungen die Rede ist. Aus diesem Grund spricht man hier von ‚fuktionalen' Beschreibungen. Da Kausalbeziehungen sowohl auf der alltagssprachlichen wie auf der mikrophysikalischen Ebene eine wichtige Rolle spielen, bieten sie die Möglichkeit, eine Verbindung zwischen den beiden Ebenen herzustellen." (2002, 16)

[4] Chalmers and Jackson in their version of the explanatory gap argument refer to the material implication $P \rightarrow M$ whose truth, as they argue, can be known "with justification independent of experience". (C&J 2001, 316). Levine himself writes that "explanation is supposed to involve a deductive relation between explanans and explanandum." (1993, 131)

process, it matters under what representation a fact is being explained. So more precisely, my question is: Does the theory of H_2O explain the fact that water boils at 100°C under the representation 'water boils at 100°C' or under the representation 'H_2O boils at 100°C'?[5] I will use "H_2O explains the boiling point of water" as shorthand for "H_2O explains the boiling point of water under the representation 'water boils at 100°C'" and *mutatis mutandis* for "the boiling point of H_2O". Of course Levine wants to explain the boiling point of water, not that of H_2O. The problem is this: To explain the boiling point of water, we have to already know or presuppose that water is H_2O. But we justify our claim that water is H_2O on the basis that it can explain the boiling point (and other macro-properties) of water. It seems that we are stuck in an explanatory circle. Or are we?

The Boiling Point of Water

Let us now look in detail at the new paradigm case: the reductive explanation of the boiling point of water. I will present the case in its standard interpretation, i.e. as it figures in the conceptual version of the argument. The reductive explanation of the boiling point of water is the explanation of the explanandum

(W) Water boils at 100°C

in terms of the chemical theory of water, i.e. the explanans

(CT) H2O at kinetic energy *e* displays molecular behavior *b*.[6]

For (CT) to explain (W) means for (W) to be deducible from (CT). A rough sketch of the deduction would look like this:

(*a priori reduction*)

(CT) H_2O at kinetic energy *e* displays molecular behavior *b*.
(1) The causal role of water is to boil at 100°C, freeze at 0°C, be liquid at room temperature, etc. [conceptual knowledge]
(2) H_2O boils at 100°C, freezes at 0°C, is liquid at room temperature, etc. [from (CT)]
(3) H_2O fills the causal role of water. [from (1) and (2)]
(4) The substance which fills the causal role of water is water. [conceptual knowledge]

Therefore

(id) H_2O is water. [from (3) and (4)]

Once we have deduced the identity claim, we can then deduce

(W) Water boils at 100°C. [from (2) and (id)]

It looks as though with a little help of conceptual knowledge we were able to evade circularity. Let us go through the deduction step by step.

Stage I: Conceptual Analysis

Two of the above premises, (1) and (4), are to be justified by conceptual analysis, i.e. they depend on conceptual knowledge. While (1) spells out the causal role of water and is thus about water, (4) makes the claim that our concept of water is of that substance which fills that causal role; (4) is thus about our concept of water. Let us see whether the analysis of our concept of water actually yields these premises.

---

[5] Levine 1998 (453) similarly distinguishes the conceptual possibility of a situation *S* under representation *R*, which may differ from the conceptual possibility of *S* under *R'*.

[6] In premise CT *e* should be taken as an independent variable. For any value of *e* CT tells us what behavior *b* the H2O molecules exhibit.

It is perplexing to note that Levine at no point offers an explicit analysis of the term 'water'. The most he says in favor of (1) is that there are "various superficial properties by which we identify water – its liquidity at room temperature, its freezing and boiling points etc." (1993, 131) But note that other superficial properties by which we identify water are not properties that specify the causal role of water, e.g. water comes out of the tabs, falls from the sky, and fills the lakes and oceans. It is commonly assumed in the discussion that these properties among others are the ones by which we identify water. However, they do not specify a causal role. So, the causal role of water, known by conceptual analysis, will comprise something like the following properties: liquid at room temperature, boils when heated, freezes when cooled, potable, thirst-quenching, odorless, transparent, etc.

Some critics have argued that we cannot give a precise definition of the causal role of water simply on the basis of conceptual knowledge. This challenge can be assuaged by modifying (1) from an identity claim into a conditional.

(1') If something boils at 100°C, freezes at 0°C, is liquid at room temperature, etc., then it fills the causal role of water.

The claim that (1) can be known simply by having the concept of water seems too strong a claim. It would imply that we know the precise and complete causal role of water simply by having the concept of water. (1') is less demanding. It only implies that, if someone who has the concept of water were presented with a complete and precise description of the causal role of water, she could identify it as such.

Let us turn to the more controversial premise (4) about the concept of water. (4) says that 'the substance which actually fills the causal role of water is water.' This can be split up into several claims.

(i) *Causal role.* Our concept of water is exhausted by the causal role of water. This means that whatever fills the causal role of water, is water.

(ii) *Acquaintance.* Water is what *actually* fills the causal role of water. Since Kripke's *Naming and Necessity* it is a commonplace that the concept of water is a rigid designator, denoting the same thing in every possible world. (Kripke 1980, 128) This means that on twin earth where the watery stuff is not H_2O but *XYZ*, what twearthlings call water is not what we call water; water ain't twater. (cf. Putnam 1975) Thus, when we want to know the referent of our concept of water, we need to make sure we are looking at the liquid, odorless, potable stuff around here, and not, say, on twin earth.

(iii) *Natural kind.* Another idea sparked by Kripke is that 'water' is a natural kind term. This idea is reflected in (4) by the term 'substance'. I take this claim to mean that water is individuated by its microphysical structure. Someone may erroneously read the above acquaintance-condition to say that H_2O on Mars is not water, since we are not acquainted with it, i.e. we are not acquainted with tokens of Mars-H_2O. But as this condition specifies, what matters is the type, the micro-physical structure. Since we are acquainted with the structure H_2O being water, H_2O on Mars is also water.

All three claims can be found in the following passage on the analysis of 'water'.

> I think we have to recognize an *a priori* element in our justification [of the identification of water with H_2O]. That is, what justifies us in basing the identification of water with H_2O on the causal responsibility of H_2O for the typical behaviour of water is the fact that *our very concept of water is of a substance that plays such-and-such a causal role* [(i) causal role condition]. To adopt Kripke's terminology, we might say that our pretheoretic concept of water is characterizable in terms of a 'reference-fixing' description that roughly carves out a causal role. When we find the *structure* [(iii) natural kind condition] that *in this world* [(ii)

> acquaintance condition] occupies that role, then we have the referent of our concept. (1993, 131; my emphases)

(iv) *Uniqueness*. One last complication may arise. What if it had turned out that there are different kinds of watery stuff around here? (Cf. Gertler 2002, 32-37) What our concept of water would refer to in such a case depends heavily on how the scenario is exactly spelled out, e.g. on how different the kinds are, how much of one kind is present on earth relative to the other kind, what kind of acquaintance-relations we have to each kind, etc. In any case, for our purposes of reductive explanation we can again simply consider a conditional claiming that if there is one unique (or very dominant) kind that fills the causal role of water, then that kind is water. This yields

(4') If there is one unique (or very dominant) kind of our acquaintance which actually fills the causal role of water, then that kind is water.

A great deal of controversy surrounds the question of what exactly we can know about water based on conceptual analysis. However, the two conditionals developed above evade many of the critiques and are sufficient for the deduction from (CT) to (W). Note however, that the strategy to use conditionals instead of identity claims is not explicitly mentioned by Levine. However, he stresses that the reductive explanations require bottom-up necessitation only. Construing the deduction as using conditionals simply reflects this directedness of epistemic necessitation.

Stage II: Empirical Description

The empirical description of a boiling water situation is the explanans. In my sketch of the deduction there are two empirical premises, (CT) and (2). While (CT) can be taken to be true ex hypothesis (the question is, whether we can deduce (W) from (CT)), we may wonder how to derive (2), the claim that H_2O boils at 100°C, freezes at 0°C, is liquid at room temperature, etc. from chemical theory. Let us first take a look at Levine's microphysical description.

> Molecules of H_2O move about at various speeds. Some fast-moving molecules that happen to be near the surface of the liquid have sufficient kinetic energy to escape the intermolecular attractive forces that keep the liquid intact. These molecules enter the atmosphere. That's evaporation. The precise value of the intermolecular attractive forces of H_2O molecules determines the vapour pressure of liquid masses of H_2O, the pressure exerted by molecules attempting to escape into saturated air. As the average kinetic energy of the molecules increases, so does the vapour pressure. When the vapour pressure reaches the point where it is equal to atmospheric pressure, large bubbles form within the liquid and burst forth at the liquid's surface. The water boils. (Levine 1993, 129)

Of course, Levine admits that this story is incomplete, that it is but an idea of how a reductive explanation does work in fact. The description is still full of macro terms such as 'surface', 'liquid', 'atmosphere', 'bubbles bursting' etc. which, in a serious reduction, would have to be translated into micro terms. Still, Levine concludes the above quote by saying that "it is inconceivable that H_2O should not boil at 212°F at sea level" (1993, 129) How do we deduce statements involving terms such as 'boiling' and '100°C' from statements involving 'intermolecular attractive forces' and 'kinetic energy'? This is the problem of bridge principles which is interconnected with the empirical description. Since the bridge principles are the crucial part of the explanatory gap argument, the problems with it cannot be appreciated unless one understands the problems surrounding the bridge principles. Therefore, I will allow myself to digress on this issue and subordinate the exegetical issues for the remainder of this section.

Bridge Principles[7]

Consider again the vocabulary of the above quote. We have seen that the empirical description still contains a lot of macro-terms like 'surface', 'liquid', 'atmosphere' etc. The lack of a smooth transition from micro- to macro-description becomes apparent in the last two sentences: "When the vapour pressure reaches the point where it is equal to atmospheric pressure, large bubbles form within the liquid and burst forth at the liquid's surface. The water boils." How de we know that the liquid is water? How do we know that bubbles bursting is boiling? And how do we know that there are bubbles bursting to begin with? Strictly speaking, all this doesn't follow from chemical theory. However, Levine in concluding the passage quoted above claims that from chemical theory we can derive the statement (H_2O) 'H_2O boils at 212°F'. Why can't we directly derive (W) 'water boils at 100°C' from (CT)? Because from chemical theory alone does not follow the identity of water with H_2O. But does it follow from chemical theory that kinetic energy *E* is 100°C or that molecular behavior *B* is boiling? No, it doesn't. Just like we need 'H_2O is water' as a bridge premise in the reductive explanation of the boiling point of water, we need bridge premises for 'boil', 'liquid', etc.

Generally bridge principles are statements of the form $Fx \rightarrow Gx$[8], where F is specified in a different terminology than G (e.g. micro vs. macro, part vs. whole, or scientific vs. folk). In the story outlined above, they facilitate the translation of the microphysical description given in terms of 'motion of H_2O molecules', 'breaking bonds', 'atmospheric pressure', etc. into an everyday macro-physical language, i.e. into the words 'water boils at 100°C'.

> The problem is that chemical theory and folk theory don't have an identical vocabulary, so somewhere one is going to have to introduce bridge principles. For instance, suppose I want to explain why water boils, or freezes, at the temperatures it does. In order to get an explanation of these facts, we need a definition of 'boiling' and 'freezing' that brings these terms into the proprietary vocabularies of the theories appealed to in the explanation. (1993, 131)

So, according to Levine, which bridge principles do we need for the reductive explanation of the *boiling* point of water—a bridge principle for 'boiling' or also one for 'freezing'? The above quote is equivocal, but best read as making the claim: For the explanation boiling we need a definition[9] of 'boiling', for the explanation of freezing we need a definition of 'freezing'. However, that claim is not quite right. It might be true if we substituted 'H_2O' for 'water', i.e. if we were to explain the boiling or freezing point of H_2O. But, to explain why water boils, we have to first deduce that water is H_2O. And to deduce that water is H_2O, we have to show that H_2O plays the causal role of water. That is, we have to show that H_2O boils at 100°C, freezes at 0°C, etc. So, to derive the identity statement 'H_2O is water' we need bridge principles of 'boiling', 'freezing', and all the other macro-properties of water. It seems as though Levine underestimates how many bridge principles are needed just for the reductive explanation of the boiling point of water. I admit that this point is somewhat pedantic, as the quoted passage doesn't imply that we need the definitions of boiling or freezing *only*. Still, I like to stress the importance of the bridge principles whenever I can.

---

[7] The term 'bridge principles' originated from the discussion on theory reduction in the 1960s. Even though the bridge principles needed for reductive explanations are somewhat similar to the bridge principles in theory reduction, they should be kept apart. In some cases theory reduction is concerned with phenomena that do not bear on reductive explanations, such as diachronic reduction of earlier theories by more recent ones. I will use the terms 'bridge premise' and 'bridge principle' synonymously.

[8] Of course bridge principles can be biconditionals as well, but usually a simple conditional will be sufficient as we are concerned with 'bottom-up necessitation' only.

[9] "Definition" here should be read to mean a definition in microphysical terms, so that such a definition actually figures as a bridge principle.

Another issue concerns the epistemic status of the bridge principles. While Levine 1993 acknowledges the need for translation and bridge principles he does not tell us in which of the two stages of reduction the translation is located. Is it rather conceptual work (stage 1) or empirical work (stage 2)? As the conceptual argument hangs on the question whether there is an a priori entailment from explanans to explanandum, I take the bridge principles to be a crucial part of the argument and I propose to treat the translation as a third stage of its own. This is not a substantial critique of Levine and others, but a matter of emphasis.[10]

Even though it is widely accepted that bridge principles are needed, they are rarely stated. One may indeed wonder why no one ever seems to give it a try: Is it just a painstakingly complex matter or is it impossible, after all, to give an a priori definition of liquidity in microphysical terms? As far as the bridge principles are stated or otherwise articulated, it becomes clear that different authors have different things in mind, hence my criticism of neglect. One question that leads to considerable disagreement is whether the bridge principles are a priori or a posteriori.[11] To answer that question let us take a closer look at the bridge principles that are needed for the reductive explanation of the boiling point of water. To deduce (W) 'Water boils at 100°C' from (CT) 'H_2O at kinetic energy *e* displays molecular behavior *b*' we primarily need bridge principles that connect 'water' with 'H_2O', 'boiling' with 'molecular behavior *b*' and '100°C' with a 'kinetic energy *e*'. One way to connect 'water' and 'H_2O' is via the identity claim 'water=H_2O'. Generally, for reductive explanations a conditional like 'if something is H_2O then it is water' would do. But Levine considers identity statements only, presumably because he is attacking the identity claim 'pain=Cff'.

However that may be, the conceptual argument showed how (id) could be deduced from (CT). Yet, in the deduction we had to presuppose that in the step from (CT) to (2) bridge principles for concepts such as 'boiling', 'liquid', '100°C' etc. are available. These we shall turn to now. Levine tells us that "the obvious way to obtain the requisite bridge principles is to provide theoretical reductions of these properties as well."(131) He further writes that "the justification for this reduction will, like the reduction of water to H_2O, have to be justified on grounds of explanatory enrichment as well." (132)

First, what are we to make of the talk of "theoretical reductions"? Are they the same as reductive explanations? One way to understand 'theoretic reduction' would be this. Identity claims such as 'molecular behavior b is boiling' can either figure as theoretic reductions or be the conclusion of reductive explanations. They are theoretic reductions insofar as they figure as premises in other reductive explanations and are justified by explanatory power. They are reductive explanations in so far as they figure as the conclusion in a deduction. This interpretation isn't too helpful, but must do for now. I will come back to the question what may be meant by 'theoretic reduction' in the last section.

Now, the reductions of the macro-properties are the bridge principles we are looking for. They principally look like this:

(boil) Molecular behavior *B* is boiling.
(temp) Molecular kinetic energy *E* is 100°C.

Given (boil) and (temp) we can deduce from

[10] To see that the bridge principles are being neglected, see e.g. Dempsey (2004) where in a two-page discussion of Levine's explanatory gap argument the bridge principles are mentioned only in a footnote.
[11] See for example Block and Stalnaker: "All we reject is the a priori, purely conceptual status attributed to the bridge principles connecting the ordinary description of the phenomena to be explained with its description in the language of Science." (1999, 8f.)

(CT')[12] H_2O at kinetic energy *E* displays molecular behavior *B*

the desired premise

(H_2O) H_2O boils at 100°C.[13]

These sketches of bridge principles (boil) and (temp) are just meant to illustrate in what way they figure in the deduction. Obviously we cannot know very much about their epistemic status as I just stated them using *E* and *B* as placeholders for actual values. Levine tells us that they are justified by their explanatory power, that is, they are justified because they figure in successful reductive explanations. Note for now that whether there is an a priori entailment from explanans to explanandum essentially depends on the epistemic status of the bridge principles. Only if the bridge principles are a priori can we claim an a priori deduction. In their present form they hardly look like it.

But now, what do we make of the fact that that the very macro-properties of water, before they can be explained, have to be "reduced" themselves? Are we back in the explanatory circle? This depends on how the reduction of the properties in question is done, or in other words, how the required bridge premises for 'boiling', 'freezing', 'liquidity' etc. are justified. If these bridge principles can be derived in an a priori way, the reductive explanation of water is not threatened by circularity. If however, the bridge principles of the relevant properties are justified in an empirical way, i.e. because they license reductive explanations such as that of the boiling point of water, we seem to be stuck in a circle.

In the remainder of this section I will discuss the a priori reduction of the macro properties of water. That is, I will ask how the bridge principles connecting the macro properties of water to fundamental properties may be justified a priori. One of the very few authors who not only spells out the required bridge premises, but also tries to make plausible their a priori status is Ansgar Beckermann. Concerning the reductive explanation of the liquidity of water at room temperature he first identifies the functional role of liquids:

> [L]iquids differ from gases in that their volume is (almost) incompressible. They differ from solids in that their shape is changeable and moulds itself to the receptacle holding them. This provides us with an—albeit incomplete—list of the features that characterize the property of being liquid. (2000, 53)

He then proposes the following bridge premises that are required for a microphysical description of water to entail that it is a liquid. (P4) connects a molecular state to incompressibility, (P5) a molecular state to the property of adapting to a receptacle:

> (P4) If the mean distance between the molecules of some substance can be reduced only by great pressure, then the volume of that substance can be reduced only by great pressure.
>
> (P5) If the molecules of some substance can freely roll over one another, then the shape of this substance is flexible and moulds itself to the shape of the receptacle in which the substance is placed. (2000, 53)

These bridge principles are formulated as conditionals, where a state described in microphysical terms (distance between molecules, molecules rolling over one another)[14] implies a state

[12] (CT') can be taken to be an instance of (CT), where the value *E* is inserted for the independent variable *e* and *B* is inserted for the dependent variable *b*.

[13] Strictly speaking, all that follows is that at 100°C H_2O boils, and not that the unique boiling point of water is at 100°C. However, in the same way in which we have shown that H_2O boils at 100°C we could show that H_2O does not boil at any other temperature significantly below or above 100°C. Since this only complicates the argument I will not go into it.

[14] Actually, 'pressure' in the antecedent of (P4) isn't exactly micro-vocabulary. This shows how hard it is to define a macro-state in exclusively micro-terms.

described in macro-physical terms (volume of substance, shape of substance). In analogy we could try to spell out the bridge principle needed to entail the state of boiling from a micro-physical description. But first we would have to identify the property of boiling with a causal role. This is a complicated matter as it is not very clear, what is implicit in the folk concept of boiling. Some candidate features may be the following: a liquid that boils is rapidly changing its state of aggregate from liquid to gas, or simply, boiling involves rapid evaporation, boiling typically involves bubbling, etc. An amendment of (boil) might look like this:

(boil') If the molecules of some substance have sufficient kinetic energy to rapidly escape into the air, if the vapor pressure exceeds the atmospheric pressure, and so on …, then this substance is boiling.

This is still hopelessly unclear. Can we assume that something similar to Beckermann's (P4) and (P5) is available for boiling? And if so, would that be a priori knowledge? Beckermann argues that the "bridge principles actually used in science rather seem to be unproblematic a priori conditionals" (2009, 156, fn. 5). He provides the following examples of putatively a priori conditionals:

> (disc) "If all parts of a disc revolve around a certain point, the disc itself rotates around this point" (ibid),
>
> (dissolve) "An object dissolves in a liquid if its parts (molecules) are untied from each other and distributed among the molecules of the liquid" (ibid),
>
> (transp) "An object is transparent if light rays pass through it." (ibid)

These statements have a strong a priori feel to be sure. But on reflection we may actually find that our assent to them is based not on purely conceptual knowledge, but on knowledge of concepts which are impregnated with empirical knowledge. For example concerning (transp), do we really know a priori that transparency means letting light rays pass through? Doesn't the very fact that we think of light in terms of rays reflect empirical knowledge? What we certainly do seem to know from the armchair is that something is transparent iff one can *see through it*. But to know that one can see through things when they let light pass through is arguably an a posteriori matter. Concerning (dissolve) we encounter problems regarding the meaning of theoretical terms. Can assent to (dissolve) be a priori when understanding of the concepts involved ('molecules') implies knowledge of what the world is like? Similar problems arise for (boil). However, (disc) actually does seem to be an a priori truth. One thing to learn from this is that there may simply be different epistemic statuses for different bridge principles. Maybe some are a priori and others not. I will not try to solve this issue here.

The decisive question is this: If there is an a priori entailment from explanans to explanandum, this must rely on a priori bridge principles. But how can bridge principles such as (boil) or (temp) be a priori? Levine in his 1983 and 1993 has little to say on this issue. Yet, this is crucial. If there are no a priori bridge principles, there is no a priori entailment from explanans to explanandum and the whole process of reductive explanation is threatened by circularity. It must be this problem and other critiques of the conceptual argument that have lead Levine to substantially revise the argument in his subsequent writings. The new version of the argument can be understood by asking: If there are no a priori bridge premises, could the explanatory gap argument be held up using a posteriori bridge premises?

1998, 1999, 2001: The Empirical Argument

Without making it explicit, Levine in his later writings introduces a different version of the explanatory gap argument: the empirical argument.[15] This differs from the conceptual argument in the following respects: First, conceptual analysis no longer figures in reductive explanations. In its place Levine now uses empirical premises that are justified by explanatory power. Second, the difference between explanatory identity claims such as 'water= H2O' and not fully explanatory ones such as 'pain=Cff' is not that the former can be deduced from more basic theories while the latter can't, but rather that the former doesn't allow for further requests, while the latter does. In Levine's terms: The latter is gappy while the former is not. Most notably this means that reductive explanation does not require an a priori entailment from explanans to explanandum. According to the empirical argument, the reductive explanation of the boiling point of water looks something like this:

(*empirical reduction*)

(CT') H_2O at kinetic energy *E* displays behavior *B.* [from (CT)]
(id) H_2O is water. [justified by explanatory power]
(temp) Kinetic energy *E* is 100°C. [justified by explanatory power]
(boil) Molecular behavior *B* is boiling. [justified by explanatory power]

(W) Water boils at 100°C

Note that this deduction contains empirical premises only. While all the bridge premises, (id), (temp), and (boil), are the same premises as in (*a priori reduction*), they figure as a priori analyses in (*a priori reduction*), and as empirical identity claims in (*empirical reduction*). Levine boldly tells us that "[t]here is no analysis of our concept of water underlying our acceptance of its identity with H_2O. We accept it because of its explanatory power." (1999, 9) But how do we justify the other required bridge premises? Levine tells us that they all are justified by explanatory power. This means that even though I sketched the reductive explanation of the boiling point of water as a deduction, the whole undertaking of reductive explanation is much more of an inference to the best explanation than a straightforward deduction. I will come back to this issue shortly.

The empirical version of the explanatory gap argument can be understood to be a response to several criticisms of the conceptual argument. First, it is a reply to the objection that the conceptual analyses required for a priori reduction are not available. Obviously (*empirical reduction*) doesn't rely on conceptual analysis. Criticism of this kind has been put forward in Block and Stalnaker 1999.[16] It should be noted that Block and Stalnaker attribute to Levine 1993 the conceptual version of the explanatory gap argument. (cf. 1999, 2ff.) Their objection is that there are no explicit definitions of water stating necessary and sufficient conditions to be had a priori. Instead they draw a picture of reductive explanation that closely resembles the one used by Levine in the empirical reduction:

> Why do we suppose that heat = molecular kinetic energy? Consider the explanation given above of why heating water makes it boil. […] Identities allow a transfer of explanatory and causal force not allowed by mere correlations. Assuming that heat = mke, that pressure = molecular momentum transfer, etc., allows us to explain facts that we could not otherwise

[15] The central themes are first expressed in his 1998 paper "Conceivability and the Metaphysics of Mind". They appear again in a short paper based on a conference speech "Conceivability, Identity, and the Explanatory Gap" (Levine 1999) and are expounded in a lot more detail in his 2001 monograph *Purple Haze*.
[16] Levine already refers to this paper in his 1998, when it was yet to be published.

> explain. Thus, we are justified by the principle of inference to the best explanation in inferring that these identities are true. (Block and Stalnaker 1999, 23f.)

I take it that this is basically the same account as the empirical reduction of Levine's empirical argument. While Levine speaks of "explanatory power", Block and Stalnaker allude to simplicity considerations and "inference to the best explanation".

Second, Levine's later writings seem to be fueled by the idea that identities need no explaining. This claim has been urged also by Block and Stalnaker 1999, but also by Papineau (1998) and others. The idea is that there is no point in asking why *X*=*Y*. Levine (1998) points out that when we question identity statements, of course we do not ask why it is that something is identical with itself. This indeed would be ridiculous. Rather, when we ask why *X*=*Y* we can be doing either of two things: First, we may be asking for evidence for the identity statement. Second, it may be the case that one and the same thing is referred to by two distinct properties (e.g. Venus can be referred by 'the evening star' or by 'the morning star', where 'evening star' carries with it certain connotations, different from the ones carried by 'morning star'). We can then ask how it is that one and the same thing displays these two distinct properties. In the case of water the corresponding question is, how something that has the surface properties of water can have the micro-properties of H_2O. The reductive explanation goes to show that having the micro-properties of H_2O amounts to the very surface properties of water.

Levine writes:

> That water is H_2O is not the conclusion of any derivation. Rather, it functions as a premise in various explanatory arguments which have descriptions of water's macro properties as their conclusion. […] If asked for an explanation of [the identity of H_2O and water] the correct response is to express perplexity about what it means to explain an identity anyway. Things are what they are; there is no sense in explaining that. What we might be asking for is evidence for the identity […] But the evidential question is answered by pointing to explanations of other facts, such as the fact that water is liquid at room temperature, which depend crucially on acceptance of the identity of water and H_2O. (Levine 1998, 462)

I assume that what Levine claims for the bridge identity (id) is equally true for the other bridge premises.

The empirical argument is an adequate response to the two criticisms described above. However, two new objections arise immediately: First, aren't we now back in an explanatory circle? Second, in what sense is the explanandum epistemologically necessitated by the explanans? If it is not, which seems to be the case, where is now the disanalogy from 'water=H_2O' to 'pain=CFF'?

I think the first objection can be defused. Is the explanatory framework circular? Even though I sketched (*empirical reduction*) as a deduction, the premises cannot be justified independently of the conclusion. The sense in which the premises are not justifiable independently of the conclusion is the following. In all the reductive explanations of the macro-properties of water there will figure the premise (id), 'water is H_2O'. Therefore, the explanation of all the macro-properties of water such as its boiling point, its freezing point, its liquidity at room temperature, etc., depend on the truth of (id), i.e. our justification of (W), 'water boils at 100°C', etc., depends on the truth of (id). But, remember how Levine proposes to justify the truth of (id). It is justified *because* it figures in those reductive explanations just sketched. It is justified because of the explanatory power it conveys when it figures as a premise in reductive explanations. This justification can best be understood as an inference to the best explanation. To sum up, (id) cannot be justified independently of the reductive explanations in which it figures. Rather, with the chemical theory of H_2O the whole explanatory apparatus comes into place simultaneously. So, in a sense the explanatory framework is circular indeed, but that doesn't really mean that it's defective. I think, we have to take this as a peculiarity of scientific theories. They are not built

step by step but function as large networks of interdependent statements. The interdependency brings with it a sort of circularity, yet not of a vicious kind.

The second objection, however, poses a serious problem for Levine. Since (*empirical reduction*) relies solely on empirically discovered bridge identities, a similar deduction might be possible from 'CFF' to 'pain'. All we need to find is the relevant bridge identity of the type 'CFF=pain' and we're done. Levine is aware of this problem and answers that there remains still a fundamental difference between 'water=H_2O' and 'CFF=pain' – the latter is gappy whereas the former is not. Let me elaborate.

Levine argues that „the reason we don't find a gap in [the reductive explanation of the boiling-point of water] has nothing to do with the availability of an analysis of 'water'". (1999, 8) He claims that the identity-statement 'water=H_2O' is simply a 'nongappy identity', which means that it doesn't allow for explanation: "An identity claim is 'gappy' if it admits of an intelligible request for explanation, and 'nongappy' otherwise." (Levine 1999, 8) A corresponding psycho-physical bridge identity, say 'pain=CFF', would however be gappy, i.e. it would allow for intelligible requests for explanation. I have no quarrel with Levine's classification into gappy and nongappy identity-statements. In fact, it takes us back to the very beginning of the argument, viz. to the distinction between fully explanatory and not fully explanatory identity statements.

However, simply claiming that 'H_2O=water' does not allow for explanation whereas 'pain=CFF' does, will not convince the materialist. Without a further argument of why one explanation is gappy and the other is not the opponent of the explanatory gap argument will ask: Why is it that, as you claim, 'water=H_2O' doesn't admit of further inquiry whereas 'pain=CFF' does?[17] I believe that there are two answers to this question.

First, the reason why 'water=H_2O' is nongappy whereas 'pain=CFF' is gappy lies in the very fact that there is an analytic conditional of the kind "if something is H_2O then it is watery stuff", and an analytic conditional of the kind "if something is the unique watery stuff around here, then it is water". It is *because* water admits of a conceptual analysis in terms of its causal role that the identity claim water=H_2O is nongappy. It is precisely because phenomenal states *do not* admit of an analysis in terms of causal role, that identity statements between phenomenal and physical states must remain gappy.[18]

Second, the difference between the identity claims may be due to a difference in the semantics of the predicates 'water' and 'pain'. Both 'pain' and 'water' are rigid designators. But while 'water' works rather like a name, 'pain' works partly like a description with an analyzable content. I am drawing here on a distinction made by Beckermann (2009, 167ff.). Predicates which work like names admit of different questions than predicates that have an analyzable content. With respect to name-like predicates science answers what-questions of the type: What is the essential nature of *X*? For example, science tells us that the essential nature of water is H_2O. There is no further sense in asking why water is H_2O. On the contrary, with respect to analyzable predicates, a paradigm case being water-solubility, science does not answer what-questions, but why-questions. We do not ask science, what the essential nature of being water-soluble is. For we know this simply by knowing the meaning of 'water-soluble': *X* is water-

[17] To be fair: Levine in his (1998) and (2001) does offer an argument in favor of this claim. However, his argument is rather complex and the discussion of it beyond the scope of this paper.

[18] This is contrary to what Levine in his 1999 states (even though not in his other papers). There he writes: "The problem is that (iii) [the bridge premise: Qualitative state R is the state that plays causal role C] isn't analytic. While it may be true that experiences with a certain reddish qualitative character tend to play a certain causal or functional role, it doesn't seem to be a conceptual truth that they do." (1999, 6) I believe that the problem with psychophysical bridge-premises is not that they are not analytic, but that they can't be stated in terms of a causal role at all.

soluble iff *X* dissolves when immersed in water. Rather we ask, why is it that, say, salt is water-soluble. Science then explains why salt is water-soluble by referring to the molecular structure of salt and water and how they imply the water-solubility of salt. The upshot is this, only with respect to predicates that admit of an analysis does it make sense to ask for explanations. 'Pain', as Beckermann argues, is more like 'water-solubility' than like 'water'. Thus, 'water=H_2O' needs no explaining whereas 'pain=CFF' does.

1993 Reinterpreted

In this last section I will argue that the standard interpretation of *On Leaving Out* is not beyond doubt. I will attack the standard interpretation in claiming that the basic ideas which later lead to the formulation of the empirical argument are already present in 1993. Of course I may be criticized of projecting Levine's later thoughts onto his earlier writings. But I think that this criticism can be met as there is plenty of textual evidence for these thoughts in his 1993. In fact, the whole paper contains a great tension in swaying between the conceptual and the empirical version of the argument. Thus, I will not claim that 1993 is best interpreted as expressing the empirical argument. Rather, I will claim that it sways between the empirical and the conceptual argument.

Let me begin by briefly rehearsing the main difference between the conceptual and the empirical argument. According to the conceptual argument there is an a priori entailment from the chemical theory of H_2O and conceptual knowledge of water to (W). According to the empirical argument (W) cannot simply be deduced from chemical theory. Rather, it follows from several scientific hypotheses, such as (id), (boil), and (temp), which are justified by their explanatory power. Therefore, there is no a priori entailment of (W) from chemical theory and conceptual knowledge.

To show the relevance of my claim I will now present some quotes to prove that Levine 1993 is generally interpreted to express the conceptual argument, i.e. I will show that there is a standard interpretation according to which Levine's idea of reductive explanation requires an a priori entailment. For example, Liam Dempsey explains Levine's account of reductive explanation as a two-stage process. He illustrates the account with the explanation of the boiling point of water:

> Consider again the reduction of water to H_2O. First we determine the causal role of our pretheoretic concept of water. [Dempsey of course means to say that we determine the causal role of water, not that of our concept of water.] Once it is established that H_2O fills the causal role that defines water, we can say that water=H_2O. In addition, according to Levine, if water can be analyzed in terms of its causal role, then *its behavior can be deduced from facts about the chemical theory* of the properties that fill that role. In other words, facts about the chemical theory of H_2O *necessitate* facts about the macro properties of water. For example, the fact that, at sea level, water boils at 100°C *is deducible* from the chemical theory of H_2O. (2004, 227; my emphasis, Dempsey's emphasis omitted)

Karol Polcyn in summarizing Levine's explanatory gap argument writes the following:

> [E]xplanatory relations are conceptual in the sense that the phenomenon to be explained is *entailed a priori* by the explaining phenomenon. Consider again the identity "Water = H2O" and suppose we want to explain why water boils at 212ºF. […] The key point is that assuming that the explanatory mechanism is in place (and assuming that the physical and chemical laws are as they are) *it will follow a priori that water should boil at 212ºF*. (Polcyn 2005, 52; my emphasis)

Janet Levin in her (2002) discusses the question whether reductive explanations require conceptual analyses. In introducing the challenge she writes:

> More recently, Frank Jackson, David Chalmers, and Joseph Levine have re-emphasized the need for [an intelligible] connection, suggesting, in particular, that the qualitative character of an experience must be *deducible from*, or *logically necessitated by*, facts about the physical process in question. (Levin 2002, 571; Levin's emphasis)

Ausonio Marras begins his 'Consciousness and Reduction' with the following claim.

> A number of philosophers—among them Joseph Levine […]—have claimed that there are conceptual grounds sufficient for ruling out the possibility of a reductive explanation of consciousness. Their claim assumes a functional model of reduction […] which requires an *a priori entailment* from facts in the reduction base to the phenomena to be explained. (2005, 335; my emphasis)

Or, as I have already mentioned, Block and Stalnaker take Levine to profess the conceptual argument:

> Levine, Jackson, and Chalmers suppose that the gap between descriptions in terms of microphysics and descriptions in terms of, for example, 'water' and 'heat' is *filled by conceptual analysis*. (Block and Stalnaker 1999, 23; my emphasis)

And:

> [W]e are not persuaded by Levine's argument, …. All we reject is *the a priori, purely conceptual status attributed to the bridge principles* connecting the ordinary description of the phenomena to be explained with its description in the language of science. (Block and Stalnaker 1999, 8f.; my emphasis)

In direct reply to Block and Stalnaker Chalmers and Jackson claim, somewhat more cautiously, that a "close tie between reductive explanation and a priori entailment is suggested by Chalmers 1996, Levine 1993, and Loar 1997." (Chalmers and Jackson 2001, 32, fn. 26)

Finally, consider Beckermann's interpretation of Levine's explanatory gap argument. We have seen above that he tries to provide an a priori reduction for the bridge principles in the reductive explanation of the liquidity of water. He introduces that passage by saying "Let us try to unpack what Levine has in mind here by means of the example of the macro-property of being liquid." (Beckermann 2000, 53) This means that on Beckermann's interpretation of Levine, reductive explanation requires a priori bridge principles. Another point is this. Whether the liquidity of water has been explained, according to Beckermann, hangs on the following question: "Can we now derive the fact that water is liquid at a temperature of 20° C from the properties of its molecules?" (Beckermann 2000, 53) So, again, Beckermann interprets Levine as looking for (a priori) deductions. In his 2008 he writes:

> Unsere Informationen über die physische Welt mögen so vollständig sein wie möglich, wir können Levine zufolge aus ihnen nicht *a priori* ableiten, welche Wesen Schmerzen haben und welche Wesen sich freuen (2008, 434)

The fact that we cannot deduce a priori the mental states from the physical states simply means that we cannot reductively explain the mental states. So, reductive explanation, according to this interpretation, requires a priori entailment. I think these quotes sufficiently show that indeed there is a standard interpretation of *On Leaving Out*, according to which *On Leaving Out* expresses the conceptual argument, i.e. the argument which understands reductive explanation to require an a priori entailment from explanans to explanandum.

I will now contrast the textual evidence for both interpretations. In doing so I will concentrate on the question of exegesis, not on the substantial question about which interpretation yields the better argument. What speaks in favor of the conceptual interpretation has already been portrayed in the second section. I will briefly rehearse: The first point concerns the epistemic

relation between explanans and explanandum. Levine says that in reductive explanations "the phenomenon to be reduced is epistemologically necessitated by the reducing phenomenon." (1993, 129) Since only on the conceptual version the explanandum can be deduced from explanans using solely logic and conceptual knowledge, only on this account can the explanans be said to epistemologically necessitate the explanandum in a sense strong enough for the explanatory gap argument. The second point concerns conceptual analysis. According to Levine "explanatory reduction is, in a way, a two-stage process" that involves on the one hand "the (relatively? quasi?) *a priori* process of working the concept of the property to be reduced 'into shape' for reduction" and on the other "the empirical work of discovering [the] underlying mechanisms". (1993, 132; Levine's emphasis) The conceptual interpretation nicely fits this two-stage account of reductive explanation. It involves both a priori conceptual analysis and empirical research. The empirical interpretation, as we have seen, doesn't involve conceptual analysis at all.

Let us now look at the points which speak in favor of the empirical interpretation. First, consider Levine's reference to explanatory power. The conceptual argument does not seem to fit with the idea that the identity-claim of water with H_2O is justified by explanatory power. On this account, (id) follows from our conceptual knowledge and the chemical theory of H_2O. There is simply no need to rely on justification by explanatory power, which would amount to something like an inference to the best explanation. Only on the empirical interpretation do we need to rely on justification by explanatory power. It is more than one passage in which Levine explicitly states that (id) as well as the bridge principles such as (temp) and (boil) are justified by explanatory power.

As a similar point, consider the following:

> [T]he identification [of water with H_2O] affords a deeper understanding of what water is by explaining its behaviour. (1993, 131)

In saying that the "identification" of water with H_2O explains the behavior of water Levine seems to take the identification not as the conclusion, but rather as an assumption from which explanations flow. Even the very fact that he speaks of an "identification", i.e. he literally speaks of an *act of identifying A* with *B*, and not, say, simply of an identity statement, portrays that the identity statement is actively assumed rather than concluded upon. We have seen that he later professes this exact position.

As a third, but related point, remember the question as to what was meant by 'theoretical reductions' as opposed to 'reductive explanations'. On the conceptual interpretation it is hard to make out a difference between the two concepts. Yet, Levine's use of 'theoretical reduction' strongly suggests that he has something in mind different from reductive explanation. He uses the concept in the context of justification by explanatory power. On the empirical interpretation this allows for a persuasive interpretation: Theoretical reductions are identity-claims that are justified by explanatory power in the context of a specific theory. For example, on this account we can say that with the reduction of water to H_2O Levine refers to the identity-claim of water with H_2O which in turn explains the boiling point of water:

> What is explained by the theory that water is H_2O? Well, as an instance of something that's explained by the reduction of water to H_2O, let's take its boiling point at sea level. (1993, 129)

The same goes for bridge principles. Levine tells us that "the obvious way to obtain the requisite bridge principles is to provide *theoretical reductions* of these properties as well." (1993, 131) And further that "the justification for this reduction will, like the reduction of water to H_2O, have to be justified on grounds of explanatory enrichment as well." (1993, 132)

Finally, consider a point concerning the explanation of identities. It may seem as though Levine's shift to the empirical argument has been strongly influenced by the surge of papers claiming that identities need no explaining. (He definitely does react to Block and Stalnaker 1999.) However, it is noteworthy that even this idea has been present in his writing from the very start. For example, in his (1993) he never speaks of the explanation of the identity claim 'water=H_2O', but only of the explanation of properties. More interesting, even earlier, in his (1983), we can find precursors of this idea. First, remember that the whole argument begins with the assumption that some identity-claims are fully explanatory (and hence need no further explaining), whereas others are not. Second, in spelling out this difference he asks for "rules which determine when the demand for further intelligibility is inappropriate." (1983, 358) This passage strongly resembles Levine's later distinction between gappy and non-gappy identity-claims. And, in that sense, the empirical argument isn't that new at all; it rather spells out some ideas that have been present from the start.

Conclusion and Outlook

My aim was to examine how the reductive explanation of the boiling point of water figures in Levine's explanatory gap argument. I have shown that there is a tension in Levine's classic formulation of the explanatory gap, i.e. in his (1993). It is unclear whether he embraces the conceptual or the empirical version of the argument, thus leaving open the question whether reductive explanation requires an a priori entailment from micro- to macro-statements. I have also argued that the answer to this question essentially depends on the epistemic status of the bridge principles. For this reason they deserve a lot more attention than they get in the present debate of the explanatory gap argument.

The conceptual version is wrong in supposing that the truths of certain conditionals were knowable independent of any empirical experience. For, the relevant concepts figuring in the conditionals are impregnated with empirical experience, even if of a very basic kind, that was essential in shaping the concepts in the first place. The empirical argument fails for the opposite reason, namely for supposing that conceptual analysis didn't matter at all. Without reference to the semantics of 'pain' and 'water' it cannot be shown that there is a relevant difference between the allegedly gappy 'pain=CFF' and the allegedly nongappy 'water=H_2O'. The next step in the evolution of the explanatory gap argument could be to try to combine the strengths of the two versions into one new argument. For this purpose we would have to address the question of the epistemic status of the bridge principles. There seem to be bridge principles that are arguably analytic, as e.g. Beckermann's (disc) quoted above. Others, such as (transp), seem to require some very basic empirical knowledge (viz.: 'if an object let's light rays pass through, one can see through it'). Yet others, the ones involving theoretical concepts, e.g. (boil), seem to be true only in virtue of our empirical theory. To get a better hold on this issue we would need to examine the following question: Given that scientific terms are part of a large interdependent conceptual framework, what does it mean for a statement containing theoretic terms to be a priori or analytic?

Bibliography

Beckermann, Ansgar (2000): The Perennial Problem of the Reductive Explainability of Phenomenal Consciousness – C. D. Broad on the Explanatory Gap. In: T. Metzinger (ed.): Neural Correlates of Consciousness – Empirical and Conceptual Questions. Cambridge MA: MIT-Press, 41-55.

-- (2008): Analytische Einführung in die Philosophie des Geistes. Dritte Auflage. Berlin: de Gruyter.

-- (2009): What is Property Physicalism? In: Ansgar Beckermann, Brian McLaughlin, Sven Walter: The Oxford Handbook of Philosophy of Mind. Oxford: Oxford University Press, 152-172.

Block, Ned & Stalnaker, Robert (1999): Conceptual Analysis, Dualism, and the Explanatory Gap. Philosophical Review Vol. 108, No. 1, 1-46.

Chalmers, David J (1996): The Conscious Mind. Oxford: Oxford University Press

Chalmers, David J. & Jackson, Frank (2001): Conceptual Analysis and Reductive Explanation. The Philosophical Review, Vol. 110, No. 3, 315-360.

Dempsey, Liam P. (2004): Conscious experience, reduction and identity: many explanatory gaps, one solution. Philosophical Psychology, 17:2, 225-245.

Jackson, Frank (1998): From Metaphysics to Ethics. Oxford: Oxford University Press.

Kripke, Saul A. (1980): Naming and Necessity. Cambridge MA: Harvard University Press.

Levin, Janet (2002): Is Conceptual Analysis Needed for the Reduction of Qualitative States? In: Philosophy and Phenomenological Research, Vol. 64, No.3, 571-591.

Levine, Joseph (1983): Materialism and Qualia: The Explanatory Gap. Pacific Philosophical Quarterly 64, 354-361.

-- (1993): On Leaving Out What it's Like. In Davies, M. and Humphreys, G. (eds.) Consciousness. Oxford: Blackwell Publishers, 121-136.

-- (1998): Conceivability and the Metaphysics of Mind. Noûs 32:4, 449-480.

-- (1999): Conceivability, Identity, and the Explanatory Gap. In: Stuart Hameroff, Alfred Kaszniak, David Chalmers (eds.): Toward a Science of Consciousness III. Cambridge MA: MIT Press, 3-12.

-- (2001): Purple Haze. Oxford: Oxford University Press.

Marras, Ausonio (2205): Consciousness and Reduction. In: British Journal of the Philosophy of Science, 56, 335-361.

Nagel, Thomas (1974): What Is it Like to Be a Bat? In: The Philosophical Review 83, 435-50.

Papineau, David (1998): 'Mind the Gap'. In: J. Tomberlin (ed.), Philosophical Perspectives 12: Language, Mind, and Ontology, Oxford: Blackwell, 373-88.

Pauen, Michael (2002): Einleitung. In: Michael Pauen / Achim Stephan (Hrsg.): Phänomenales Bewusstsein – Rückkehr zur Identitätstheorie? Paderborn: Mentis, 9-34.

Polcyn, Karol (2005): Phenomenal Consciousness and the Explanatory Gap. In: Diametros, Nr. 6, 49-69.