Types and Tokens in Folk- and Neuropsychology - A Philosophical Study of Psychological Taxonomy

Masterthesis
in Philosophie
at the University of Konstanz

by
Carsten Griesel

Submitted:
24 November 2006

Supervisors: Professor Dr. Wolfgang Spohn
Professor Dr. Harald Schupp

Index of Contents

Types and Tokens in Folk- and Neuropsychology - A philosophical Study of Psychological Taxonomy 1

I Abstract 3

II Preface 3

III Introduction 6

IV What exactly is the Identity Theory of Mind? 11

IV.1 What motivates the Identity Theory? 11

IV.2 Arguments against Identity Theories 15

IV.2.a First Argument against the Identity Theory: The “Explanatory Gap” 16

IV.2.b Second Argument against the Identity Theory: Kripke’s necessity Argument 18

IV.2.c Third Argument against the Identity Theory: The argument of “multiple realizability” 20

IV.3 Why Token Identity is perceived as uninteresting 22

V The difference between Types and Tokens 23

V.1 The Relation between Types and Tokens 24

V.2 Type Generation exemplarily demonstrated in Biology 28

V.3 Identity Criteria for Token and Types 32

VI Natural Kinds of Psychology and Neurology 36

VI.1 Problems common to Psychological and Neurological Type Generation 37

VI.2 Candidates for Neurological Types 39

VI.3 Considerations about Psychological Types 47

VII Implications for the Reliance in Introspection 50

VIII Conclusion: Scientific Taxonomy results in Type Identity if Token Identity holds 51

IX References 54

I Abstract

The thesis that predicates about the mind refer to the same entities as predicates about the body is investigated under the focus of the type/token distinction. Furthermore, it is analysed, what scientific types – the categories of a scientific taxonomy – are. After objections to the identity thesis are considered, it is argued that the philosophical view about folk psychological types stands in contrast to scientific types which are the target of psychophysical identity claims. Finally, the analysis of types being the result of categorising tokens with respect to their micro structure is used to elaborate the claim that type identity holds if token identity does.

II Preface

This thesis is a defence and endorsement of a theory claiming that psychological phenomena are completely reducible to neurological phenomena[1] which means that psychology and neurology refer to the same objects (i.e. have the same extension). The so called "identity theory" became popular in the late 1950s/early 1960s with the appearance of three influential publications (Feigl 1967; Place 1956; Smart 1959). Thus this thesis, together with the topic, is situated right at the junction of philosophy, neurology and psychology – in order to use neurological findings in psychology, one simply needs a theory of the relation between the brain and the mind!

[1] It has become somewhat out of vogue to engage in such a project. This is mainly because there is some tension between reductionism and the status of the special sciences. As a result, people favour speaking of a „reductive explanation" (Kim 2005, pp. 93ff). I see the point as denying that a macro phenomenon is actually the same as some micro structures underlying certain physical laws. This would result in a very narrow understanding of identity. On the other side, it seems to me a matter of taste how narrow one wants to understand identity in this context. After all, it seems possible to deny that H_2O is actually identical with water. Since this would seem to me somewhat arbitrary, I keep calling a reductive explanation of psychological phenomena in the way that these phenomena are completely explainable by the underlying micro structures with the respective natural laws, an identity.

All three disciplines are core parts of probably every cognitive science institute in existence. Still, there is an essential problem between philosophy and the other two disciplines: The questions as well as the proposed answers formulated by a philosopher do not at all seem to satisfy psychologists or neurologists and the same happens frequently in the inverse case. Being unsatisfied wouldn't be so bad because it would mean that one just has to refine and reformulate questions and answers so long as the situation improves. But the actual situation seems to be even worse for most interactions between philosophers and psychologists as some of them even claim that the work of the one has nothing to contribute to the work of the other.

Specifically philosophers advancing the argument of multiple realizability (which will be discussed in section IV.2.c in greater detail), claim that neuroscience has little to contribute to the understanding of psychological processes.[2] It is noteworthy that this argument, together with two others, was mainly responsible for identity theory being driven out of the ongoing mind-brain debate for quite some time. On the other side, psychologists and neurologists often doubt that the writings and the very far reaching questions of philosophers have any implications on their work. This is due to the fact that part of the philosophical work is often speculative. As for now, the technical possibilities to answer the "big question" for the complete mind-brain picture are not yet given and so it is not surprising to see an empirical scientist denying a possible contribution of a philosopher's work. Thus, the intuitive picture of philosophers asking the questions and the psychologists going out and trying to answer them (e.g. van Gelder 1998) does not seem to hold. Still, psychology is – as most other sciences are – a spin off of philosophy and the disciplines divided because (as with the other disciplines) it became apparent that it is impossible to answer the questions asked by philosophers without empirical work.

[2] Viz, such philosophers see the problem between the *other* two disciplines.

As a quasi by-product of this thesis, there should be an answer to what a psychologist can learn from a philosopher and vice versa. It will be shown that falsification of a philosophical hypothesis (like the identity thesis of mind) depends on the interaction of conceptual and empirical work. So, there is a work-task from the philosopher to the experimenter. It might be possible that philosophical considerations alone show that a certain theory *cannot* be formulated without inner contradictions. Such considerations can be a guide for the experimenter. This has to be said with caution, since philosophers tend to proclaim the inconsistency of theories at an early stage. Because of this, it seems advisable not to give too much attention to the "worthlessness claims" concerning neuro-empirical work. Still, and without doubting the ability of neurologists and psychologists to fit their findings in a general context, philosophers can evaluate the importance of these findings in an even bigger context, since philosophers formulated the initiating interest in the influence on the idea of men of a certain scientific domain. For the highly detailed answers of today's special science though, it is not always obvious how they impact on the idea of men. This is still the philosophers' work and for this, speculations and counterfactuals definitely have their value.

On the other side, this thesis is written to explicitly object to the philosophical trend declaring some empirical findings as *necessarily* impossible. It is not that I disbelieve in the power of logical deduction, yet I am convinced that some philosophers were far too quick with such an extensive claim such as the worthlessness of neurological findings to our understanding of the mental.[3] By sketching a plan for generating (strict) psychophysical laws, this thesis naturally aims towards showing that the hunt for such laws is still as fruitful as it can be.

[3] I suppose that one reason for the willingness to state a worthlessness claim is the use of oversimplified examples. See footnote 8 for an explanation.

III Introduction

At some time in their life, most people with a cultural background like Christianity, where a conceptual line between an immaterial soul and a material world is drawn, wonder how the relation between their mind and everything else in the world (their body, the environment, etc.) can be understood. Historically, people who could not desist from this question, engaged in philosophy. As the method of introspection had proven to be insufficient for further investigation of "the mind", psychology split off from philosophy and empirical work was also up to the understanding of mental ongoings.

It seems to be an empirical fact that for most people in the western hemisphere, dualist intuitions are a starting point for their considerations. This is probably due to the fact that mental activities are given in a first person perspective while all other observations are accomplished from the "outside".[4] This obviously suggests a fundamental difference between psychological phenomena and physical ones. But in today's physicalized world, scientifically minded people quickly get the idea that it is not compatible to a scientific understanding of the world to postulate something that corrupts numerous of well established natural laws. Additionally, today's progress in psychology and neurology make a very strong case (if not already a proof) that thought and behaviour have at least something to do with the brain. Thus, it seems unavoidable to find a way for integrating mental experiences of living beings into the existing body of well established natural laws. A rather straightforward method to carry out such a unification of the mental and the physical is to claim that these two ontological categories (or "res" as Descartes has called it in the Meditations on first philosophy, Descartes 1986) belong in fact to the same ontological category meaning that a mental state or event is nothing else than a physical state or event.[5] This was actually achieved

[4] There are other reasons for the impression of the special status of the mental as well but this seems to be the most inevitable.

[5] A lot of trouble arises when labelling "nothing else as" as identity. Of course it is a stronger claim for two things to be the same than claiming that one thing is nothing else than another which means that it is reducible to

in philosophy with the already mentioned "Identity Theory" which is the topic of this thesis. Unfortunately this was initially done in such a straightforward way that it was inevitably false. The standard example for establishing an identity between mental and physical states was to propose that pain[6] – representative for other mental conditions – is actually nothing else than the firings of *certain kinds* of neurons.[7] But if this held, then it should be possible to extract such a neuron (or more of them, it would not matter) from a creature, put it into a patch clamp apparatus (Kandel, Schwartz, and Jessell 2000, p. 111), turn it on and voila, you would have pain lying on your experimentation-table.[8] From this we learn that such a thing as C-fibre activity can at best be a necessary condition for pain but it definitely is not the sufficient condition for it and thus is far from being identical with pain.

Consequently, people started looking for a less specific relation between neurons and mental states or events. The next step was to propose not an identity between some mental states or events and the firings of certain neurons but to propose an identity to the complete "brain states" at the time of the mental states or events. And indeed, if we would "extract" the whole brain from a creature, and would be able to put this brain in the same state as if it

---

the other. One could ask, if water is really the same as H_2O or even if the property of being water is the same property of being H_20. (For an interesting discussion, see Abbott 1997.) In the morningstar/eveningstar example it is certainly not the case. Here everybody agrees that these terms denote the same object (i.e. the planet Venus) but of course it is a different property to shine bright in the evening than to shine bright in the morning. In the case of water one could probably say that a lot of H_2O molecules together do have the same properties as water. With Venus this does not work, because here we have no reduction of the two terms.

[6] Here and through this thesis, only the psychological component of pain is meant. It is unequivocally be true, that pain has also a behavioural component but for our purposes, the seemingly unphysical things that go on besides the observable behaviour is of interest. Pure behaviourism is thus discarded as not capturing the whole story.

[7] It was popular at the time of the first versions of the identety theory to think that "C-fibre" firings are potential candidates for being identical with pain. Discovered at that time, activity in these nerves is still thought to be correlated especially with kinds of chronic pain.

[8] It is a nuisance that philosophers are so stuck to use the C-fibre example all the time. Though they are right in stressing that it is only a placeholder for something that is a better candidate for an identity relation, this placeholder is so oversimplified that it has from the beginning only little plausibility. Unfortunately the property of being implausible seems to be passed from the placeholder to everything it represents.

would be still "embodied" and its bearer would feel pain, it sounds much more plausible to assume that the extracted brain instantiates pain. But evil minded philosophers did not allow the answer-seeking people to settle with the "brain-state-explanation" once and for all. The main problem with this kind of reductive explanation was an argument which will show up in section IV.2.c under the label of the "multiple realizability argument". In brief, the objection is the following: If we undertake a reductive explanation of e.g. pain in the way that we identify pain with a brain state, say Brian's brain state at time *t* – the time he was in pain – then it is extremely unlikely that anybody or anything will ever instantiate pain again. This is obviously so, because it would be a galactic coincidence to ever see anything again that has exactly the same brain with the same neuronal activation pattern as Brian's brain at time *t*. The natural response to this is that an identification of a mental type like pain with an exact brain state was far too pretentious. Mental vocabulary (it does not matter whether we speak of folk psychology or scientific psychology here) consists of classificatory terms – it is part of any psychological endeavour to taxonomise the entities of a theory of the mental. Terms like "pain" denote a mental type because a lot of sensations have something in common that suggests grouping them together under the term "pain".[9] This means that a lot of (slightly) different instantiations (or the properties that are essential for them) constitute the type-term "pain" (and all other mental terms respectively). In the philosophical discussion, these instantiations are called "tokens".

In this thesis my main concern is to clarify the relations between *types* and *tokens* in the context of mental and neurological states or events. Following my current line of thought,

[9] Most prominently, their functional role, but for individuals who classify their own feelings, the nature of the feeling does certainly play an important role too. This seems to be overseen often and might be a consequence of the "Private Language Argument" from Wittgenstein (Wittgenstein and Anscombe 2001). Besides from the argument, I do think that there is no need for a neural network like the brain to linguistically classify e.g. an emotion. It is quite debatable that the brain monitors its own activity in any way. But if it does, this would be a possibility to classify certain brain states. This shows that there is at least no principle reason why psychological events should not be classified on the basis of introspection.

it is almost too obvious to declare that the brain states – as philosophers perceive them – should be regarded as *neurological tokens*. But this view has not been very popular. The reasons for this will be discussed in greater detail in the next chapter. The motivation for declaring brain states not as neurological tokens but as neurological types was grounded in the curiosity to learn something about mental phenomena like pain in general. Mental types were of interest and not something about a certain mental state or event. So theorists just *needed* a neurological *type* that could be identified with a mental *type*, otherwise only a mental *token* could be identified with entities on the neurological level. But as should be clear, much will be won already if it is admitted that it makes sense as well on the mental side as on the neurological side to differentiate between types and tokens.

This leads us to the insight that nobody ever claimed that every pain feels for every person (always) the same, but merely that pain feelings have something in common.[10] Now, it is a much more realistic claim, that every single pain is identical to a single brain-state which then is a "token-identity-theory". And even this can be further extenuated because a large part of someone's brain activity is probably unimportant for his pain feeling. Eventually being in pain does not mean that my mental life consists of pain only. Even in the moment of pain sensations, there are other mental activities going on. I still can think and feel a lot of things while sensing pain. So we probably would be only claiming an identity to the *relevant* neural activation or "parts of brain states".

Now we are at the heart of the topic of this thesis. What actually happened in the search for the correct mind-brain relation was that some philosophers indeed claimed that "only" an identity between psychological *tokens* (single states or events) and neurological tokens existed (Davidson 1970, 1973, 1974; Kim 1966). This was of course due to the

[10] It could also be that all pain feelings only stand in a "family resemblance" relation to each other (Wittgenstein and Anscombe 2001). This is an interesting idea which is not even unlikely. Since this would complicate the matter at this time, it is ignored for the moment. But it may become a topic again if empirical work shows complications in "defining" mental states.

obvious failure of the claimed identity between a psychological type and a neurological "brain state" that was also supposed to be a type but can in fact only be a token – a concrete instance and not an abstraction defined by a prototype or a set of essential properties. After all, a brain state is firstly a concrete "state" and not a state of a certain type. So, proposing a "token identity" is actually a much more modest claim.

But why did this claim not have a bigger impact on the philosophical discussion? There were actually two reasons for this. First, many people were of the opinion that token identity claims were far *too* modest to be of interest. It just would not give an answer to the question people were interested in, namely, how mental phenomena such as pain could be explained. At best, it would give an answer to how *a* pain could be explained. The second reason for denying the scientific value of the token identity claim – which is quite interrelated to the first reason – is that it is obviously thought that type and token identity theories are somehow competing in the way that one could not assume any kind of type identity theory as there are only chances to make a token identity plausible (for the view that a token identity does not constitute a type identity, refer to Beckermann 2001, p. 140; Fodor 1974).

I will argue for both reasons to be completely misleading. As already indicated, I whish to specify how psychological types and tokens as well as neurological types and tokens can be understood. Hereby the neurological side is certainly more interesting and debatable. However, to complete my picture, something has to be said about psychological types as well. I do not think that this step will be in any way revolutionary but it still has to be explained why my view about psychological type-terms and their origin stands in some disagreement with the classical philosophical use. This will connect psychological types with psychological tokens. I will also draw a picture about neurological tokens and types and how they are related. Certainly, the biggest quest in this is to find a plausible view on "neurological types". But if there is a plausible account for neurological types, it will become clear that the identity of neurological and psychological tokens means that the types can also be "harmonized" in

the way that even type identity holds. By bringing mental and neurological types in accordance, it will be shown that the manner, in which the types become specified, is what is informative about such mental phenomenon as pain. This stands in contrast to the classical view that assumes the pure existence of an identity is what is most informative.

IV What exactly is the Identity Theory of Mind?

As already mentioned, the identity claim is quite an exacting claim. Many philosophers seem to believe, that it is too strong to hold. In this Chapter I will first explain why – in the end – there is no alternative to a reductionist explanation, such that the theory of interest is a possible version of an identity theory instead of some fallback position making weaker and weaker claims. I also have to explain the problems of identity theories in greater detail and why it fell into disgrace. After that, I will substantiate my picture of the type/token distinction and warrant it with the help of examples from other disciplines.

IV.1 What motivates the Identity Theory?

The answer to the headline's question is fairly simple: The need for explanatory power. What we are up to, is an explanation for the properties of mental phenomena (why one psychological state is often followed by certain others, why it has the qualities that is has for us and even why we have so much trouble grasping all of this, etc.). The naïve understanding of mental phenomena, the so called "folk psychology", gives only a very rough idea why mental phenomena have the properties they have. However the major problem with folk psychology is not its inaccuracy but its being unintegrated into the body of the existing natural laws. So, as long as psychology is not in any way connected to what else we know about the world, it will have to be regarded as mysterious. Consequently, we do not only need a theory with explanatory power but it has to be compatible with the well established theories of physics as well. Causality and causal closure of physics are immense problems for an autonomous theory of the mental which finally made dualism unattractive. The alternative is

"physicalism"[11], which in turn means the possibility to connect entities of the to be explained phenomena to some established theory by reduction, i.e. to say that a macro phenomenon "consists in" the (complex) interplay of micro structures with certain properties and the macro phenomena follow the laws they do because of the laws which underlie the micro structures.[12,13]

The possibility of reductive explanations is exactly the strength of the identity theory. Take the following example. A teacher orders one of her students to the blackboard to write down the result of a mathematical exercise she just posed. To calculate the result is probably one of the most prototypical mental engagements. What happens here is that the student hears the request to come to the blackboard which means that sound waves are transduced to neural signals by the eardrum, ossicle and the cochlea, and then something has to happen with these neural signals finally causing the student to get up and to write the result on the blackboard.[14] What I just addressed with "then something has to happen" is normally described as a mental process and from a physical point of view it is at best a black box description. However, I have already delivered a hint for the "interface" between a standard physical description and mental processes: the transduction into neural activity. With regard to motor output, we also

[11] The opposite of "dualism" is not necessarily "physicalism" but monism. There is not only materialist monism i.e. physicalism but since I regard the well established body of knowledge of the natural sciences as starting point for further considerations, a mental monism seems a very unattractive position.

[12] Philosophers who interpret the identity relation too tight and therefore deny the possibility of such an identification (in the domain of the mind body problem at least), are nevertheless interested in some kind of physicalism or materialism. In a 1989 article Jaegwon Kim (Kim 1989) is critical about the prospects of such an endeavour. In his 2005 book "Physicalism, or something near enough" (Kim 2005), he seems to have changed his mind with the argument that "[t]here is no consensus on exactly how nonreductive physicalism is to be formulated, for the simple reason that there is no consensous about either how physicalism is to be formulated or how we should understand reduction." (p. 33) I guess he changed his mind because he became sceptical if his view (supervenience) could count as reductive. The point is that weaker versions of materialism are not exactly incompatible to physical laws but they lack what is interesting about an explanation. I hope this will become clear with the rest of this chapter.

[13] So understood, reduction thus tries to deduce complex laws from simple laws as in axiomatic mathematics.

[14] A similar example can be found in Beckermann (2001).

know that this is done by neural activation of the muscles. In addition to today's knowledge that even in-between there *is* neural activity going on, it seems to be very strange to propose that the neural signals produced by the cochlea are again transduced into "unphysical" signals that must be translated again into physical signals when it comes to produce some behaviour. The only open question from a purely physical perspective is whether or not our knowledge about the laws underpinning neural activity suffices to explain what it was that had to happen. If this can be done, then this is exactly the ascertainment for this certain mental calculation of the result being the same process as the neural activity that happened "in-between" and was relevant for the performance – just described at the micro level. Only that we have "de-black-boxed" what we labelled as mental before and could claim that we have understood it.

It certainly makes a difference if we regard psychology (scientific or folk) as a theory and now want to connect this theory to another theory (to explain events) or if we "simply" want to connect the objects of the different theories (the states of affairs). Of course, the difference is that in the first case, we also have to connect "rules" to other rules. This is obviously the more interesting thing to do and Kim (2005, p. 107f) proposes three possibilities to do so:

(i) *Bridge laws, or trans-ordinal laws – contingent, empirical laws connecting explanandum phenomena with phenomena at the reduction base.*

(ii) *Conceptual connections, e.g. definitions, providing conceptual/semantic relations between the phenomena at the two levels.*

(iii) *Identity statements that identify the explanandum phenomena with certain lower level phenomena.*[15]

[15] Kim himself actually favours the second option for very good reasons. His reasons are too good to just ignore them. So, I obviously have to answer the question why I still speak of identifications. The reason will hopefully be clear at the end of this thesis but to interpret my project: It will look like the line between option two and three will vanish or that possibility (ii) has to be done even if one goes the third way but the connections have to be

The difference between these three possibilities is the "closeness" between explanandum – here, mental events and their properties – and the explanans – here, a micro-level explanation in the form of neuroscience. There is one philosophical position that denies the possibility of any of these three ways, namely dualism. Consequently, for dualism, mental events have to stay mysterious. That would mean that we would never understand *why* mental states of affairs have the properties they have and *why* mental events underlie the rules they do. Because this would be such a mayor exception to scientific principals, it is the very last option to be considered. So, while dualism is the complete decoupling of the mental from everything else (thus we cannot even speak of explanans and explanandum anymore), identity is the closest coupling of explanans and explanandum.[16] In between, proposing "some coupling" of the mental and the physical, there are proposals for the relationship between mind and body like emergentism, epiphenomenalism, functionalism and supervenience. In this listing there is also some gradient of closeness with supervenience being the closest relation between mind and body next to identity. Emergentism has several versions which differ in the closeness between the micro and macro phenomena where the stronger versions which state that the macro phenomena are irreducible, new and their structure cannot be predicted by the micro level description also have to be counted as dualistic (Stephan 1999). For functionalism and supervenience one can descry that there is a relation between them so that they can be combined (Kim 2005).[17]

My point is that already supervenience (and therefore the others even more) lacks something that only identity theory can deliver. In order to briefly recapture the idea, supervenience states that mental properties supervene over material properties i.e. that there

---

done on the inner psychological domain (namely between its types and its tokens) and not so much between the psychological and the neurological domain. Here identification is favoured.

[16] Conceptual connections might be even closer when they are about *known* identities.

[17] Though, functionalism is traditionally perceived as "anti-reductionist" (Putnam 1967b; Fodor 1968). Some do connect functionalism with identity theories (especially token identity theories) instead of superveniences (Carrier and Mittelstraß 1995, p. 58; Levine 1983).

cannot be any change of a mental property that is not at the same time a change of a physical property whilst it can be the other way around. An often mentioned example for supervenience is the property of an image to depict a certain object which supervenes over the physical properties of the picture (see Lewis 1994). Thus, a supervenience relation is perfect for claiming that many different micro instantiations make up one macro phenomenon. The problem is that this does not explain why a certain macro phenomenon supervenes over exactly these micro instantiations. It just describes the looseness between explanans and explanandum but it does not deliver the explanation.[18] In other words supervenience gives no answer to how and why such mysterious mental properties supervene over certain material configurations (and not over others). Thus, the strategy of loosening the closeness claims between explanans and explanandum to deal with the supposed problems (which are depicted in the next section) had the consequence of loosing what was interesting. This has to be the case since only reduction delivers explanations of the macro phenomena with the help of the micro structures and the laws that hold for them.

IV.2 Arguments against Identity Theories

Now that we have learnt about the merits of the identity theory and why we should endorse it, I shall evaluate in greater detail which objections were made against it. The most influential arguments against identity theories were Saul Kripkes argument of the necessity of identity statements and the multiple realizability argument (most famously brought forward by Hilary Putnam and Jerry A. Fodor) as well as Joseph Levine's argument of the explanatory gap. The three arguments are connected in the way that Kripke's argument seems to emphasise the strength of the multiple realizability argument and Levine's argument builds upon Kripke's considerations.

[18] A defender of the supervenience claim would probably say that this is no problem for the claim since the real explanation for the "why" has do be delivered by the neuroscientist and not by the philosopher. Since philosophers are interested in understanding the mind-body relation, I disagree about that.

At the end of this thesis, I shall review the situation and evaluate whether these arguments really do have the capacity to make the identity between the neurological and psychological types look implausible.

IV.2.a First Argument against the Identity Theory: The "Explanatory Gap"[19]

Let's first say something about the explanatory gap since we have already come across the topic of explanation in the last section when I claimed that explanatory power is exactly one reason why we should seek a reductive explanation. The expression "explanatory gap" was exemplified by Joseph Levine in two influential articles (Levine 1983, 2002). Levine originally formulated his argument against a reductionist explanation, while Kim (2005, p. 94) notes that a reductive explanation or the reduction of some mental phenomena to some physical phenomena were often thought to close the gap. So do I.

What is in question is whether an answer can be found concerning the issue "why pain, not itch or tickle, arises out of C-fibre stimulation[20] […]" (Kim 2005, p. 94) without leaving out any open questions. But indeed, the C-fibre example makes the point of the explanatory gap obvious: Why should any understanding, any gain of knowledge arise from the simple claim that such a thing as pain is actually the same thing as C-fibre activity? It just seems arbitrary to make such a claim and as such, it definitely seems to be a contingent fact – which later will play a role in the Kripke argument. This kind of identification – given as such – entails no *explanation why* C-fibre activation should denote the same event as pain and not itch. What we would need is something about C-fibre activation that even brings predictive power about why C-fibre activation can *only* be pain and feel like it. Instead it seems as if we had no arguments – besides the a posteriori encounter of the correlation of pain and C-fibre activation – to counter someone who actually claims that C-fibre activation is in fact the same as an itch.

---

[19] My counting is on no way chronological. In fact one could say it is just the other way around.

[20] Again, the usual term that is used as a placeholder for any neurological explanation.

Levine himself (1983) saw the problem to be even bigger. He was optimistic about seeing the causal role[21] of mental phenomena like pain explained by the laws underlying the realising micro phenomena, i.e. neural activity, but he declared that this is not all that has to be done in case of mental events. The explanation of the causal role of a macro phenomenon by the causal role of the entities at the micro level would be sufficient in examples like the identification of water and H_2O and heat and the mean molecular kinetic energy. However in the case of pain and the like, it would have to be explained why "pain feels the way it does" (or has the "quale" it has) too – and not only why it has the causal role it does. Consequently, John Foster (1994, p. 301) formulates:

> "Our conception of *P* [a pain-event] in terms of its psychological (introspectively manifest) character seems to offer no clue as to how it could also have a neural character, and our conception of *N* [a neural event] in terms of its physical (scientifically discoverable) character seems to offer no clue as to how it could also have an experiential character."[22]

To me, this seems to say that it must be explained why the behaviour associated with pain is associated with it and additionally claiming the need to explain why pain feels like pain and not e.g. like an itch.[23] This need is also implied by the distinction between "p-consciousness" and non phenomenal consciousness introduced by Ned Block (1995). The "problem of consciousness" (and in the final consequence the "qualia problem") that is behind this can probably not be dissolved so easily. But what should be kept in mind is that psychological types (as all other types) are individuated by their causal role as Kim rightfully states with his

[21] Since "causal role" is the most central notion of functionalism and Levine regards it as successful reduction if the causal role of a macro phenomenon can be explained by showing how this causal role is realised, a connection between identification and functionalism can be seen here.

[22] The famous "knowledge argument" of Frank Jackson aims at the same intuitions (Jackson 1986).

[23] The "zombie debate" is a huge debate about the possibility of mental states having different properties, e.g. feel different or do not have any experiential character at all, and still have the same causal role (cf. Lenzen 1998).

"principle of causal individuation of kinds" (Kim 1992, p. 17). This implies that the causal role of mental types, for which one may have hope to explain cannot be dissociated from the "experiential character" for which Foster, Levine, Chalmers (1996) and others see little or no chance for explanation.

IV.2.b Second Argument against the Identity Theory: Kripke's necessity Argument

The second argument which attracted a lot of advertence was Kripke's argument for the necessity of the identity of objects denoted by "rigid designators", which he presented in the early seventies in a lecture held at Princeton University (Kripke 1971, 1980). Kripke's merit is that he was the first to notice that there are not only contingent a posteriori truths and necessary a priori truths, but also necessary truths a posteriori and contingent truths a priori.[24] Why is this important for an identity statement? Until Kripke, when contingently true statements were always thought to be a posteriori true statements, nobody discerned a problem in the contingency of identity statements like "pain = C-Fibre firings" – and in the last section we already saw that there seems to be a very strong intuition about the contingency of those statements: It appears not even contingent but arbitrary to claim such an identity as "pain = C-Fibre firing". It seems to be far too plausible that a certain neural activation pattern could instantiate other mental properties than it actually does in the case of oneself. Kripke underpinned this plausibility by pointing out that it could well be that one encounters being in pain not to be the same as the neural state of affairs with which one tried to identify it – the pure possibility of proposing a conclusive counterfactual is enough here. In

[24] This can most easily be demonstrated with examples. To the standard cases of contingend a posteriori truths like "Angela Merkel is the chancellor of Germany in 2006" and the necessary a priori truth like "All bachelors are unmarried" we can think of sentences using indexicals that are a priori but contingently true, like "I am here now" (Beckermann 2001, p. 132). The most interesting case for us are the a posteriori truths that are still necessary true. Under this category, identity statements with "rigid designators" on both sides of an equality sign can be found. What this means still needs to be explained but it is crucial to note that terms like pain and C-fibre activation would be rigid designators.

order to once again[25] stress the contingency of the identification from pain with a neural activation pattern is thus the first step of Kripke's argument. As yet, the a posteriori character of identity statements is uncontroversial. But Kripke's claim is that identity statements about rigid designators are – though a posteriori – necessarily true. To show this is his second step. Both steps taken together, show that the identity claims about mental states are necessarily *false* as these claims are (as shown) contingent and they can only be true if they are necessarily, i.e. in every possible world, true. So Kripke argues for the necessity of identity statement like "pain = C-fibre firing" as follows to prove the point: Rigid designators – primarily proper names – denote entities in a rigid way because they refer to properties that are *essential* for that entity. In the case of persons for example, one can easily imagine that Joseph Ratzinger might have a different hair colour or may be smaller or taller but one cannot imagine that Joseph Ratzinger could be a different person, e.g. Angela Merkel, while still being Joseph Ratzinger. Linguistically this just would not make any sense. We mean something with Joseph Ratzinger that has to be preserved through all possible worlds if we should still be able to think of him as Joseph Ratzinger. And the very same is true for Benedict XVI. So, if I claim that "Joseph Ratzinger = Benedict XVI", then this claim is either not true at all or it is necessarily true. When Joseph Ratzinger is in fact the same person as Benedict XVI, then this has to be true in all possible worlds since what is essential for Benedict XVI is then essential for Joseph Ratzinger too. Thus, a statement like "pain = C-fibre firing" is either necessarily true or not at all true. Eventually it is constitutional for pain that it has pain characteristics and it is essential for a certain neural state to be that neural state. So these are rigid designators as well.

[25] Smart himself stressed the contingency of neural to mental identity claims (Smart 1959).

IV.2.c Third Argument against the Identity Theory: The argument of "multiple realizability"

Although the last two arguments had quite an impact on the discussion about the identity theory, the deathblow finally came from the argument of "multiple realizability" which was particularly put forward by functionalists as Jerry Fodor (1987) and Hillary Putnam (1960; 1967a; 1967b; 1967c). Actually, the argument of multiple realizability did not only have an influence on the mind-body debate but on the attitude towards special sciences like psychology, sociology, etc. in general (Fodor 1974, 1991) and is thus a general argument against reductive endeavours.[26] Contrary to the two previously described arguments, this argument relies not only on analytical philosophical considerations, but on empirical facts as well. The empirical base also seems to be responsible for the argument's perspicuity. The advisement on which it is based seems almost trivial: When we try to identify mental phenomena like pain with a physical state, then it is clear that there always has to be this physical state when there is pain. But as stated in the introduction, that concrete physical state does not seem to be the only state that is correlated with pain. The physical configurations (neural configuration of the brain) that I am in when I experience pain throughout my life, are probably never twice the same and the neural states of other people and even animals doubtlessly differ from mine. The point that Fodor articulates (Fodor 1974) in order to generalise this argument is that the special sciences use terms that can be regarded as "natural kinds"[27] of the special sciences i.e. that they are generalised theoretical entities which do have their own theoretical value in the theories of these special sciences. Special sciences, Fodor

[26] As Kim (2005, p. 96) points out, this does not mean that "reductive explanation" is out of vogue.

[27] The debate about what natural kinds are is not at all settled (for a recent account, see Root 2000). The notion will stay relatively vague here too. The hands on definition that should suffice here is that natural kinds are the smallest theoretical entities to which type terms refer which occur in scientific laws. With the criterion of being the *smallest* entity it is excluded that natural kinds are disjunctions of other natural kinds. For Willard V. O. Quine, who is accredited for establishing the term "natural kind" (Quine 1969), "projectability" was essential too, where projectability means that what makes one token an instantiation of a type would make all other tokens with these properties as well instantiations of that type, thus allowing judgments about type membership by induction.

claims, primarily intend to find such generalisations and these terms generalise over physical instantiations that do not necessarily need to have something in common. Though the term in question can be regarded as natural kind of the special science, the instantiations together are not any justifiable natural kind of the science to which the special science is to be reduced. The picture Fodor draws is more that special science terms generalise over a whole bunch of physical instantiations and that a special science law of the kind

For all *x*: when x has *F*, *x* has also *F'* ($\forall x(Fx \rightarrow F'x)$

does not mean that *F'x* has to generalise over exactly the same physical instantiations as *Fx* (Beckermann 2001, p. 140). Consequently, psychological terms would generalise over many neurological states and Fodor not only doubts that the neurological states do have something in common – in the way one could think of a natural kind of neurology so to say[28]– but that psychological terms can be connected with an interesting (special science) law while there does not have to be such an interesting law-like relation between the physical instantiations. It is thus clear that the identification of mental types, which are nothing else than theoretical terms of the special science psychology with a natural kind of a more basic science, cannot be done in the simple, straightforward way i.e. identification with a concrete physical instance (e.g. a brain state). The identification of a special science term with a *disjunction* of physical instantiations would not help either given that this still does not allow the laws of special science to be reduced to physics. Fodor writes (1974, p. 109):

> "In particular, that one may not argue from 'it's a law that P brings about R' and 'it's a law that Q brings about S' to 'it's a law that P or Q brings about R or S'. (Though, of course, the argument from those premises to 'P or Q brings about R or S' simpliciter is fine.) I think, for example, that it is a law that the irradiation of green plants by sunlight causes carbohydrate synthesis, and I think that it is a law that friction causes heat, but I do not think that it is a law that (either the irradiation of green plants by sunlight or friction)

[28] Fodor explicitly doubts that a disjunction of several physical instantiations is likely to form a physical natural kind.

> causes (either carbohydrate synthesis or heat). Correspondingly, I doubt that 'is either carbohydrate synthesis or heat' is plausibly taken to be a natural kind predicate."

IV.3 Why Token Identity is perceived as uninteresting

The above three arguments – in particular the third argument – were articulated to counter in particular the type identity theory. As said before, this was based on the fact that the token identity theory was never really considered to be interesting enough though the mere possibility to reductively explain tokens was generally accepted.

Again it was Jerry Fodor with his "Special Sciences" article (Fodor 1974) who elaborated on the weaknesses of the token identity claim. He argues that the pure verifiability of a token identity neither implies materialism nor type physicalism nor reductionism and is therefore much weaker than what we normally understand as reductive materialism. To fulfil the criteria of either materialism or type physicalism or reductionism, token identity is indeed necessary, however, Fodor argues that it is never sufficient. For materialism, the lawfulness of the behaviour of the tokens has to be added to the pure token identity, for type physicalism one has to additionally assume that there are no "unphysical" properties on the macro level and for a reductive account, one has to presuppose that the realisations form a natural kind on the micro level when there is a macro level type. Though these are indeed distinct criteria that are not entailed in the pure token identity theory, the fulfilment of the second criterion would also lead to the fulfilment of the third.

After we learned that the mere verifiability of token identities does not prove a lot and is even compatible with some versions of dualism, one can ask whether the additional constraints necessary for a reductive, materialistic type identity theory cannot be derived from other insights. One can, as I will be arguing. After all, token and type identity theories are not incompatible. Ultimately, everybody agrees that if type identity holds, so does token identity. However, in addition to the above arguments against the type identity theories, there are two other concerns about reductionism in general. The advocates of those concerns thus aim to

deny even the attraction of materialistic type reduction. First, and this is one of Fodor's favourite arguments (Fodor 1974, 1991), there are exceptions to the laws of special sciences which could not be true if the basic science had no exceptions (which is generally assumed in philosophy although today's physics is far from this condition) and type identity would hold: "There is just no chance at all that the true, counter-factual supporting generalizations of, say, psychology, will turn out to hold in strictly each and every condition where their antecedents are satisfied." (Fodor 1974, p. 111) To him it is clear that exceptions to special science laws are a matter of fact and there is no arguing about special science laws being ceteris paribus laws. From this and a closely related concern, to aim at such things as type identities, is stigmatised: This concern is about special sciences being dispensable if type identity held (e.g. Baker 1987). We will come back to this concern in the next section.

While Fodor is right about the necessary, insufficient character of the token identity to reductive materialist accounts, I shall argue that the practice of scientific taxonomy reveals that there always are commonalities on the micro level that should enable us to speak of a natural kind of this level when a macro type is properly defined. In other words: We regard macro science types only as natural kinds if we were able to show that they correspond to a natural kind of a lower level science. Due to this practice, it is guaranteed that token identity results in type identity.

V The difference between Types and Tokens

To comprehend the – as I claim – short-sighted denial of the token identity's relevance, one has to clarify the notions of "type" and "token" first. After that, something can be said about the *relation* between tokens and types. This relation should be the same, independent from the level of description (in our case the two levels are psychology and neurology). By having clarified the conception of a reductive explanation of tokens and the relation between types and tokens, we can derive an understanding of what a reductive explanation of a type must look like.

V.1 The Relation between Types and Tokens

Our context is still the explanation of mental phenomena which are usually addressed by type terms. The concrete instances of these types are the tokens whose bundling results in turn in the type. The fact that special sciences are usually interested in types is of course responsible for the aim of reduction, namely a reductive explanation of the type. This results in the question for a reductive explanation being posed in a top-down manner. Because there seem to be definite convictions about how the mental types, which have to be reductively explained, really look like, the extension of the type is kept fixed while investigating the reduction base.

This is probably due to the direct "givenness" of phenomena like pain. Though I do not want to deny that people are experiencing types of pain, the *pain-type* is certainly *not* directly given, meaning that there is no *a priori* understanding of type terms like pain. As Levine (1983) reminds us (in a slightly different context), it would be very exceptional to find a *factum brutum* – something that is just there – on a macroscopic explanatory level. Special sciences like psychology deal with macroscopic descriptions and consequently we should not expect the types of psychology to be indefeasible "brute facts" – i.e. that these types just exist "out there", independently from any instantiations or definitions. Fodor (1974) told us that special sciences like psychology[29] aim at interesting generalisations that are relevant and can be regarded as "natural kinds" within that special science. A very important point can be stated here: One could intervene that "generalisations" of special sciences are not "natural kinds" but disjunctions of natural kinds that are made regarding common properties of the

[29] Of course, the mind-body problem often has "folk psychology" on the mind side and not scientific psychology. Paul Churchland (1989; 1992, chapter one) had to argue that folk psychology is actually a theory to counter his critics. I completely follow the argument of Churchland and my claims about scientific psychology can be conferred to folk psychology. After all, to say that folk psychology is not a theory to preserve it from elimination was always a bad move since this would even worse the situation for the defender of folk psychology. Though the defenders had in mind that claiming that folk psychology is not a theory would make it immune to all sorts of cultural/scientific changes it would mean that if it is *not even* a theory it is something of an even more equivocal status instead of something sacrosanct.

disjunctive natural kinds which are important for the actual scientific question. But if "types" were *such* generalisations one should not wonder that they do not reduce to *one* underlying natural kind! For the question of type reducibility, one thus *has* to refer to natural kind types!

So if e.g. pain is a natural kind of psychology, then this type of mental phenomenon is already a generalisation – a subsumption of what is essential for a token to be of this type. There is a very important point behind this insight, which is beautifully demonstrated by Michael Pauen (2000, p. 399) who refers to Robert W. Batterman (2000): "…not all causal properties which can be observed on the microphysical level are relevant for type generation on higher levels."[30, 31] This means that as an abstraction, a type constitutes a rise in description level given that it is a higher order structure which reduces information. Consequently, the type term's extension depends on the level of abstraction from the micro level properties of the tokens as the abstraction level determines how coarsely or finely grained the type will classify the tokens. Thereby the properties of a set of tokens to make up a type do not only rely on superficial properties of the tokens but also on their micro level structure. Accordingly, a detailed specification for the level of abstraction on which a certain phenomenon should to be investigated has to be given as well. Otherwise one will theorize without "context or frame of reference" as Bechtel and Mundale call it (Bechtel and Mundale 1999, p. 203). Of courses, the same level of abstraction has to be applied to the reduction base when one aims at a reductive identification.

[30] Translated from German by myself.

[31] The grasp of this insight is probably the motivation of the supervenience thesis in the mind-body debate (Kim 1994). Recognizing that not all properties on the microstructure are relevant for type membership on the macro level looks extremely as if a macro type supervenes over micro types. But this is not true. Supervenience is most often understood in the way that there do not have to be "a clear interrelationship between mental and physical properties…" (Beckermann 2001, p. 210). This is something Pauen and Batterman would deny. It is a difference between claiming that not all properties are relevant for type membership (which is true on all description levels) and denying a licit relationship between micro and macro level.

As a consequence of all this, the intuition of having a clearly defined picture about the types under investigation and then only have to decide if a certain token is an instance of that type seems to be more and more implausible. Eventually, the type under investigation might change due to new insights about the reduction base.[32]

Although this question does not seem to be discussed very often in the context of the identity theory of mind, it is an absolutely nontrivial question what a mental type is. My intention is to show that the picture of definite, sacrosanct types should be abandoned since the types of any folk science are only intuitive types which can be changed (or replaced as the Churchlands might say), by a deeper analysis.[33] This will happen when a deeper analysis influences people's opinions on what is essential for being of a certain type. As a consequence to the change of extension of a type term one could speak of elimination in the strict sense. But it is advisable to use the term elimination with caution because ignorance towards the fact that only the "old" *extension* of the term would be changed and not the phenomenon as such, played a major role in motivating the opponents of the identity theory of mind (Fodor 1974, 1981; Baker 1987; Davidson 1980). The presumption that macro phenomena "vanish" as soon as they have been reductively explained is very present in folk understanding of science too[34], however, it is nothing else than a fallacy (for an interesting discussion, see also Schwartz 1991; Cheyne 1993; c.f. Cruse 2004, p. 225). After all, it makes no less sense to use the macro type term even when reduction is successfully accomplished. No one stopped using the term water due to the knowledge that its chemical structure is H_20 and though mental types like "love" might change slightly in extension due to a taxonomy "update", as I propose, this has

[32] Actually, this process can also be regarded as a "moderate elimination" which might not be to far away from recent accounts of Patricia Smith Churchland (2002).

[33] A very similar point, with which I am very sympathetic too, is brought forward by Bechtel and McCauley (1999).

[34] An example that can frequently be observed in every day live is that many people are afraid of a demystification of the phenomenon "love" since they are afraid that it would either make it vanish or spoil in any way.

nothing to do with *eradicating* "love". At the end, though reduction might become scientifically successful at some point in the future, the macro phenomena type terms are always more economically to use and even usable as a (less precise) placeholder when one does not know the *exact* regularities determined by the underlying realisations.[35] Concerning the "feature" of allowing exceptions in special science laws one can say the following: There are exceptions because our scientific intuitions with which one started the scientific investigations were only an approximation of a *real natural kind* and the further refinement of the supposed types did probably not yet match a real natural kind either. Even Fodor can be assured that continuous encounters of exceptions will – at the end – alter the macroscopic types of the special sciences.

When I said that phenomena like pain are thought to be directly given – whatever this means – it is clear that what is directly given can only be an instance of e.g. a pain experience and not a type of pain. Why then does it seem likely that even somebody who experiences a mental state or event for the first time, will readily engage in type-talk about what he has experienced? (E.g.: Saying that this kind of sensation was unpleasant.) This is likely to be true because an abstraction or concept of a certain mental state or event can already be generated from only one instance. When we are confronted with an object of a certain type for the first time and are told that this is an object of type *x*, then we instantly have a picture of what it means to be *x* – however rough this picture might be. Though endless discussions can be held about whether this kind of abstraction, generalisation and concept generation is done around a proto*type* or in any other way (e.g. an essential feature list), the "how" is not so important for

[35] I hope that nobody disagrees about the increased possibilities of successful manipulation of the world due to knowledge of underlying regularities. E.g. somebody who knows – against popular opinion – that there is no causal relation between one and the next drawing of lots is less surprised that he does not win although many other combinations where already drawn and this stays with the same all the time. Consequently, a statistically educated person will probably less likely engage in gambling. I do not see why psychological and neurological should be an exception here. Of course, somebody who knows the underlying regularities of depression is – against popular belief – much more likely to know that pure animating will probably not be successful.

our present purpose. Rather it is important that it happens and this is the reason why obviously everybody believes to know what the types of mental phenomena are. Everybody can readily come up with an answer for what he thinks is essential for an ad hoc generated mental type, starting from only one instance of a belief, desire or sensation so that this instance falls in the ad hoc type. This is the same with mental types as with all other types. To emphasise the picture of types being post hoc entities, generated from a collection of tokens, let us review the prime example of taxonomy: The taxonomy of species.

V.2 Type Generation exemplarily demonstrated in Biology

Types are the result of taxonomy generation. It is most basic to all sciences to bring the objects of investigation of the respective science in some order (Fodor 1974, p. 101). The claim that physical realisations might not need to have interesting commonalities to make up a special science type implies that the ascertainment of such unfamiliarity of the microstructure entities would not influence the type under investigation. The claim of *interesting* cases[36] of multiple realizability is thus that types are resistant to the insight that there is no underlying "natural kind" on the microlevel. This claim seems to be especially prominent when investigating psychological types.

The view that mental types seem to be regarded as fixed seems quite surprising. Especially when one considers that in scientific psychology the generation of a taxonomy (e.g. of mental illnesses, c.f. the change of the definition of "hysteria" over time) underlies the same rules as in other special sciences (Acton and Zodda 2005; Haslam 2002).[37] Biology seems to be most suitable to demonstrate these rules. While in psychology every single mental event or state is a token, in biology every single life-form is a token for which it is interesting

[36] As interesting I regard cases where the realisations do not have something essential in common. Surely, every object is multiple realizable in that the exact atomic structure will not be exactly the same. This though, is neither the case the multiple realizability argument wants to attack nor is it something anybody would wonder about (Pauen 2000).

[37] Though it is debatable it psychiatric disorders really constitute natural kinds. While Haslam does, Zacher (2000) is sceptical.

to find a type that classifies the individual. As for other sciences, there is also a folk version of biology which served as the starting point for scientific exploration – as in the other sciences. The intuitions of folk biology are based on accessible, superficial data about the tokens of life-forms (individuals) for which a classification is sought. Naturally the most obvious properties are consulted when starting to generate a taxonomy of life-forms. Such properties might be "has legs", "swims under water", "flies", "eats plants" and so on. The properties chosen to distinguish individuals will certainly depend on the life-form tokens one has encountered so far meaning that categories conceived to be the basis for type generation are only drafted. As a consequence to such obvious discrimination criteria, it is not surprising that whales and dolphins were indeed part of the proto-taxon "fish", while penguins were not part of the proto-taxon "bird". Of course this does not mean that categories were scientifically useless, since they were the starting point of the investigation, which counters the worry of Fodor (1974, p. 113) who states that the complete commutability of special and basic science terms would lead to a "…lack [of] the appropriate theoretical apparatus for the [higher-level-science-] taxonomy of [lower-level-science-] events". However, it does not mean that the proto-taxa are in any way sacrosanct either. What happened when one of our ancestors who used the proto-taxon of "fish" realized that he wants to learn more about "fish" and then discovered that some of these "fish" have characteristics that are more similar to animals that he had classified as "non-fish" while all other "fish" share another common characteristic that the "strange fish" do not? What one might think would happen, but what has not happened, is that our ancestor thought: "Oh, there are no fish." He could also have adopted a stance popular with some of today's philosophers: While adhering to the original "fish" type one acknowledges that all animals swimming under water do not share a common characteristic *except* from swimming under water which explains their ability to stay under water. Hence our ancestor might have claimed that membership to the "fish" type is irreducible. After all,

the essential property of swimming under water is multiple realizable: with the help of lungs and with the help of gills!

Another possibility would be to invent a new sub category of "fish" which includes "fish" that have lungs, do not lay eggs, etc. This looks like a good approach when one still thinks that swimming under water is the only (or most) essential criterion for being a fish. The trouble is that the "strange fish" have characteristics that seem to be essential for other animals, while no other animals have the characteristics of the "normal fish". This favours the "strange fish" to be excluded from the "fish" category which becomes afterwards more a *fish* category and thus changes in its extension. The type "fish" has thus changed over time. This based on the fact that people have changed their opinion on what is essential for being fish on the basis of characteristics on a lower, finer detailed description level. Before there was a type meaning "animal swimming under water" and now this type does not exist anymore under the same label since further investigation of the animals swimming under water revealed that "swimming under water" does not make up a natural kind given that the animals swimming under water do group into one or more types of animals that are quite distinct. Now, the label fish denotes animals that have gills, lay eggs and live under water. Another reason for this is that the option of "fish" meaning a disjunction of fish, whales and dolphins is scientifically useless as it is extremely unlikely that this pseudo "fish" type can be used in law statements that represent any interesting generalisation in biology (except from those generalisation one can make about all life-forms under water). This is due to the reasons discussed in section IV.2.c showing that disjunctions not forming a natural kind are troublesome for law statements.[38]

[38] Fodors (1974) example is: "I think […] that it is a law that the irradiation of green plants by sunlight causes carbohydrate synthesis, and I think that it is a law that friction causes heat, but I do not think that it is a law that (either the irradiation of green plants by sunlight or friction) causes (either carbohydrate synthesis of heat). Correspondingly, I doubt that 'is either carbohydrate synthesis of heat' is plausibly taken to be a natural kind predicate."

In summary, we have seen that the primal type serves as the starting point for a scientific endeavour with the interest in learning something about “fish”. The primal type will be an abstraction of the *perceived as* essential properties of a probably small sample of tokens of life forms. These proto-taxa were solely defined by very obvious criteria. Further investigation of the “microstructure” of the primal type “fish” revealed that “fish” has two or more subcategories of which unification in one group would seem so arbitrary that one or more subcategories might be excluded from “fish” so that what is left is the today’s type *fish*. The same process can repeat itself when one investigates the microstructure even further and analyses e.g. the DNA. In biology, as in the other sciences, the possibilities to investigate even deeper microstructure organisations are most often paired with a controversy. The controversy is then about whether one should continue to categorise on higher level characteristics or should use the categories of an underlying description level. This happens when the deeper analysis reveals that actually two or more quite differently organised types have been grouped in one. Examples are cases in which two species that have evolved similar characteristics in parallel but whose genetic line split before the evolution of these characteristics are grouped in one type due to their obvious similarities but do not belong to one group according to a genetically defined taxonomy.

Now we can reply to Fodor’s claim that special science types bundle instantiations that do not have interesting commonalities on the lower level description (Fodor 1974). If this were true, it only shows a special science type is not a natural kind either and thus the discrepancy in micro level properties will also show a discrepancy in macro level properties. In other words: If there is no underlying natural kind for a special science type, meaning that this type is a disjunction of natural kinds, the macro level type will also be a disjunction of macro level natural kinds.

From this we learn that it can frequently change over time whether a certain token is part of a certain type. One might object that the original folk taxonomies have not changed but

simply do coexist with the different scientific taxonomies. And they do coexist since otherwise one could not statistically analyze on how they differ (Hunn 1975). This is the reason why I chose the obvious example of fish, in which the folk understanding of what it means to be fish changed due to scientific progress as well. Though there seems to be a huge time lag between changes in folk and scientific taxonomy, folk taxonomy is not definite either. The same holds for folk *psychology*.

V.3 Identity Criteria for Token and Types

Now that we have seen that type membership of a token is not a once-and-for-all matter, we have yet to clarify the criteria for tokens and types to be identical, respectively reducible.

It is advisable to distinguish cases of reductive identification from cases of identifications which are made on the same level of description. The latter ones are informing about type/token relations while the cases, where reductive explanation plays a role, have to deal with the issue of level overarching identifications which serve as explanations. One of the three possibilities mentioned in section IV.1 has to be chosen to evolve a level overarching explanation. I already indicated that I favour option III ("Identity statements that identify the explanandum phenomena with certain lower level phenomena"). This shall be justified in this section. Let us first investigate the examples and then move on to the three possibilities of explanation.

My first example deals with identifications on the same level. I was told that the Citroën C1, the Peugeot 107 and the Toyota Aygo are the same car with different looks. Of course, this is a case of identification on the same level because we want to identify a macro level entity with an entity on the same level – the entities to be identified with each other are all car types. If somebody is confronted with one exemplar of each of these three types and is asked to find out if they are the "same", he will probably investigate the cars on a description level lower than the level that might have "cars" as "natural" kinds. He will compare the parts

of which these cars are constructed and will find out that 80%[39] of them are the same. After that, he can report back that these three car tokens share so many commonalities that they also share the essential properties and can thus be regarded as the same.[40] This is how it works for tokens, but what about the generalised case of types? After all, the three car designations stand for (artificial) types, and not for individual cars!

For type identifications we thus have to ask ourselves what we believe to be essential for being one of these three car types. If we have seen one example of one of these types, we will have a rough idea about what constitutes this car type. It will almost certainly have something to do with its appearance. But from only one instance, we do not know how abstracted our abstraction has to be. Eventually we might believe that the colour of the only exemplar we have seen so far is essential for that type of cars. And if there were convertibles and other variants of these cars, one would even have to abstract from the shape of these cars in order to acknowledge that they still are of the type in question. Thus we are left with the question how we should determine the adequate level of abstraction to decide about type membership. Since type individuation relies on the causal properties of an entity (remember the "principle of causal individuation of kinds", Kim 1992) that are usually determined by the microstructure properties, even in cases of non reductive identification we need to look at a lower description level to be able to engender an appropriate abstraction.

In addition to talking about types, we are always forced to generate an idea about the type based on induction – no matter how many exemplars we might have encountered (c.f. Quine 1969).For types we thus have to check first, if a lot of car tokens share a good portion

[39] Eighty percent is a pure guess of mine but might well be. Probably the commonalities are even bigger.

[40] One could object that the most essential property of one of these cars is that they have the accordant badge on the front and back. With this, a brand would be a pure psychological phenomenon, a pure tool of marketing. Though a tendency can be observed that this is a result of globalised economics, historically a brand is more than an instrument of marketing and to be of a certain brand is so too. E.g. it might play a role in which factory a product is manufactured and who pays the employees. In the case of the three car types though, even this would not enable us finding a difference here.

of characteristics on a lower level description to define our type before we can ascertain whether or not this definition resembles the one that was generated from tokens presumably part of another type. Only if this ends up positively can we speak of the identity of types and we could discover that the three car types are cars of a common type.

Now we will complicate the ascertainment of "identity" even further by including of the issue of reduction. Note that even in the first example there was some need to go to a lower level description to define the type, but this had nothing to do with a reduction that aimed at explanation because car types cannot be "explained" with other car types. So when can we speak of an identity in the sense of reduction? For tokens, this will be the case when properties of the reduced token can be explained by the token's properties to which it can be reduced. In this case the "systemic properties" of the macro phenomena are a result of the complex interplay of entities on a lower level description where none of these entities features the systemic properties by itself. It is only in this sense, how *the* classical example of reductive identification, namely H_2O=water, can be understood. Eventually neither "H", nor two hydrogen atoms nor oxygen have the properties that are essential for water. It is expedient to speak of identification because one can regard the macro level type term as a name for the complex state or process of the micro level. With regard to the three possibilities that Kim gave us for a reductive explanation (bridge laws, definitions and identifications, section III) it should be clear now, why my argumentation leads to identification since that this process does not involve any bridge laws[41], nor is it a process of pure definition. Rather the macro level type term denotes the macro phenomenon as well as the complex state or process of the micro level. Definition only plays a role in ascertaining a type on the basis of its tokens and their micro level description.

[41] Fodor (1974) only discussed bridge laws as delivering the "explanatory ascent" as Kim calls it (Kim 2005), but there are at least the two other possibilities Kim mentions.

The classical example of heat is a little bit difficult here because it is arguable if it makes sense to talk about heat tokens. For the classical example of lightning the point is more intelligible. Here we have a lot of flash-tokens from which people obviously have created a lightning type on the basis of the appearances of flashes. What is nice about this example is that every flash is unique and still people are credibly able to say if something is a lightning or not.

To know the type “lightning” is to be able to credibly differentiate between flashes and non-flashes. If a person can do this, she must have generated an abstraction from the previous encountered flashes since no two are ever the same. So she has an idea of what is essential for lightning. The exact *physical realisations* of the macro description “flash” are never twice the same either. Consequently it is not possible to explain a flash by reduction with the exact realisation of another flash. To explain the *type* lightning, one thus will have to generate an abstraction of the physical realisations as well as find out the essential properties that makes electric discharge a flash (e.g. electrical discharge through an electric conductor does not result in an static electric arc while probably not even every static electric arc would be classified as “flash”). Thus, there can be no reductive explanation of a type without defining a type of the underlying realisations. Only with this micro level type are we able to explain the macro level type. In this, the case of lightning is different from the case of the cars.

At the very beginning I stated psychology and neurology to be reducible if they refer to the same objects, i.e. if the terms of the respective theories have the same extension. While this is the case for tokens when the macro level description can be regarded as a summary of the micro level entities that make up the macro phenomenon, the case concerning types is different: Here a type is equal in extension with another type if and only if the abstraction of the macro level tokens result in essential properties that can be explained by the abstraction of the micro level tokens' essential properties.

In the case of psychology and neurology the relation between psychological types neurological types, commonalities of the neurological tokens, which correspond to psychological tokens, explain why the psychological tokens make up the type they do.[42]

VI Natural Kinds of Psychology and Neurology

If one aims at such a strong claim as an identity-claim, it is absolutely necessary to first clarify, which sorts of things are candidates for being identical to which other sorts of things. The classical example for psychophysical type identities is that a psychological type called pain should be reduced to a neurological type called "C-Fibre activation". My main point in the last section was that we should not expect to have a clearly defined, incontrovertible type at hand when we use terms such as pain. This was because the extension of terms like pain depends on what we think is essential for a mental phenomenon while exactly the microstructure's revelation of a bundle of tokens that are supposed to be subsumed by that type is likely to change the notion about what is essential. Our first problem in the quest for psychophysical identities is thus the "slipperiness" of mental vocabulary. The second problem – whose solution would have repercussions for the first problem – is the misleading impression that philosophers have a *neural* type (or at least an idea of one) available that could be used on the physical side with which to identify mental terms. In today's discussion "C-Fibre activation" is – though most often used – merely a placeholder for whatever neurological type there might be. As indicated earlier (see footnote 8), I see a problem in using this placeholder since it seems to somehow shape *philosophers* imagination about how a neural type could look like. What is right about the C-Fibre example is that this type would at least be a type defined by neurological criteria, namely localisation and more important, histology – C-Fibres are supposed to be neurons that have their function because of the

[42] This is exactly the picture that Fodor (1974) wants to throw over board. Davidson would also disagree – that is why he called his version of a token identity theory *anomal* monism. The point where he and I do disagree is if the physical realisations have something interesting in common. It will be my aim I chapter VI to show that they do.

properties they have. Though localisation and histology are important differentiation criteria when “decomposing” the brain, these criteria are not very suitable candidates for constituting the kinds of types we are looking for since it is extremely unlikely that the information processed in the brain is coded (solely) by localisation and cell architecture. In order to progress in the quest for psychophysical identities, we thus have to say something about possible candidates for neurological types and by which criteria they are defined. If this can be done properly and we have a good candidate, we still should not be too surprised if these neurological types do not exactly match the mental types we expected to be there. Eventually, our intuitive taxonomy of mental types might turn out to be inaccurate such that peoples self ascription of mental states might be inaccurate. While this possibility seems to be quite absurd for some philosophers due to the supposed privileged access to ones own mental states (sometimes referred to as “non-observational”), this possibility is less spectacular for most psychologists. Despite the fact that it is not yet clear whether there are real cases in which people are mistaken about their own mental states, which are not only cases of *linguistic* confusion, I will argue why the theoretical possibility should not be surprising for anybody.

VI.1 Problems common to Psychological and Neurological Type Generation

One problem we already came across is not exactly a problem of psychological type generation in itself. The problem of finding the adequate level of abstraction is stressed because we want to identify types situated on two different description levels and therefore the problem is the possibility of a mismatch concerning the degree of the abstraction on these two levels. For mental classifications on the basis of introspection, the degree of abstraction does not seem to be an issue at all. Of course one could try to be more specific with the mental terms one uses and thus try to e.g. differentiate tooth-pain from headache every time one talks about pain. But there are certain limits to this. Though it is doubtful that a headache feels exactly the same on two different occasions, it does not seem to be possible to reduce the level of abstraction to zero so that every instance of a pain would be of its own type. There is

a good reason for this which becomes apparent when we consider other sorts of classifications done by humans. Almost all terms used by humans are at least minimally abstract in the sense that they generalize over different sensory percepts. Even when one looks at ones own pet, though it is exactly one instance of a life-form, the sensory percept is never twice the same and still we do not generate types for the view at the pet at a certain time from a certain angle. That we are actually able to be so unspecific is a consequence of the brain being a neural net whose main characteristic is the ability to generalise over a great variety of inputs. If the aim is now to find a type on the micro level with which a folk psychology type can be identified, the question arises as to how much difference we allow between the micro level tokens to count as instances of the same type. So again, we have to ask ourselves how the tokens of the different levels look like.

Though mental tokens seem to be clearly defined at first sight, they are not as clearly contoured as we might hope for. Philosophers prefer to divide (conscious) mental states into at least three subcategories: beliefs, desires and sensations. But what is *a* belief, *a* desire or *a* sensation? Is it really that one could say that my hunger lasted from 11:43:16 till 12:36:45 to the same degree? And even if it would, would this be a 53 minute and 29 second token? We should at least note that these states are not simply present or absent but give way into each other.

So far I have tried to stay neutral concerning the question whether we want to identify states with states or events with events. As Beckermann (2001, p. 124) notes, Smart (1959) and Place (1956) were so incautious as to identify mental *states* with brain *processes*.[43] Though this gave reason to object that this is a categorical error, this problem can be fixed easily in the manner that one *either* has to claim identity of events *or* states. States seem to be simpler entities and are therefore favoured in the actual discussion about psychophysical identities. On the other hand, processes or events seem to be more interesting since they entail

[43] I regard processes as a subcategory of events. For both it is essential that one state fades to another.

the transition from one state into another in the way that a law about a process like “every process *x* leads to process *y*” can be paraphrased to “whenever state *a* goes over into state *b*, state *b* will be followed by state *c*”. Hence I do not see a mayor difficulty in this state/event discussion. Additionally it seems quite strange to propose that an identity of states would not automatically mean the identity of the laws that connect the states on the two description levels. The problem of smooth transitions is also present for events and states. This problem though, is not one of the identity relation but of the tokens' taxonomy. If the macro level tokens transform smoothly into each other, the micro level tokens will as well. For the pure identity of the two levels it thus does not matter how we individuate the tokens from each other. For our taxonomy though, we can expect that there are tokens where a clear allocation to one type will be difficult since the phenomena under investigation are not naturally “tokened” but rather span a continuum.

VI.2 Candidates for Neurological Types

So far we have seen that neither the reduction base nor the phenomena for which a reductive explanation is sought-after are clearly defined. What is indisputable is that we do have intuitions about the phenomena of interest that will serve perfectly as a start for scientific understanding. But since I claim what we regard to be sufficient and necessary will alter after we are more informed about the microstructure, I will first explain the problems at the micro level.

It is one of the main points of Michael Pauen (2000) and William Bechtel and Jennifer Mundale (1999) that someone who uses the multiple realizability argument against the identity theory is mistaken with regards to the neurological entities that should be identical with the psychological entities. Since both papers refer to Batterman (2000), their critique on the neurological types proposed by the advocates of the multiple realizability argument go into the same direction: the “granularity” which is used to describe neurological types differs from the granularity used to describe psychological types and therefore the impression of a

many to one relation arises. What is meant by the difference in granularity is that scientific- as well as folk psychology use types that are abstracted to a degree that not only allows for inner individual token differences but for between individual token differences as well. This is exactly one premise of the multiple realizability argument: that mental states belong to the same type even if there are (irrelevant) differences. The justification for this claim is related to the "principle of causal individuation" (Kim 1992) in that what makes these states the "same" is that these tokens can all be instances of a type which occurs in a law such as: "Whenever someone is in the mental state of type *x*, she will do action *z*." And if this so generated type were usable in such laws which would be without exception, this would indeed be a good indicator that this type corresponds to a natural kind – especially if this type occurs in more than one invariable law. Exactly this degree of abstraction from unimportant differences is not granted when philosophers talk about brain states because philosophers suppose brain states to be exact "physical-chemical states of the brain", so Bechtel and Mundale. Thereby it is not that the electrical (physical) state or the concentration of neurotransmitters in certain brain areas (chemical state) is denied to be of essential importance for neurology and therefore for neurological type generation. The problem is just that – especially considering further neurological knowledge (e. g. the function of neurotransmitters relative to the brain area of occurrence) – this description makes it very difficult to differentiate between essential and unimportant properties regarding their taxonomy.

Though Bechtel and Mundale were absolutely right with their critique on the philosophical understanding of brain states, their proposal of what the reduction base of psychological states is, was not perfect either. This might not be that surprising considering 1999 was a time when the fMRI- enthusiasm was still on the rise. It is namely that they proposed activity in certain brain areas to be the substrate of mental phenomena and this is exactly what lesion- and fMRI- and PET studies can reveal. The framework of classifying neural activity with respect to the brain area in which it occurs delivers clear criteria for a

possible taxonomy but might fail to separate the activity regarding essential properties. This can be beautifully demonstrated with the example of cognitive neuropsychology which investigates brain lesions and their effects. Identification of brain regions with mental capabilities is achieved in cognitive neuropsychology with the help of double dissociations. A double dissociation is on hand when comparing the impact of damage to two different brain areas, the result is that damage to the first area will corrupt performance on task *A* and not task *B*, while for damage on the second area, it is just the other way around. In this way, one has a very strong indication that the first area is crucial for task *A* and the second area is crucial for task *B* while damage to the first area is specific enough not to interfere with task *B* and damage to the second area is specific enough not to interfere with task *A*.

It is clear that cognitive neuropsychology naturally tends towards an abstraction level where a difference in brain area is regarded as essential while more specific differences are regarded as unimportant. With a special resolution of two to three cubic millimetres (in practice "voxels" are two millimetres square and four to five millimetres long, Dobbs 2005) today’s imagine techniques provide a better resolution than just “brain area XY” but taking their relatively poor temporal resolution (one to two minutes for the whole brain, less than two seconds for a cross section) into account, it is clear that the focus is most often on finding activity within a certain area.

With regard to reducing mental phenomena to activity in certain brain areas, ththa5toyes 8lTJ-26..56

example. It is widely acknowledged that most accounts that tried to identify centres for specific functions were challenged by almost every new paper that investigated the same cognitive function because with most papers a new centre was brought into discussion that should be crucial for a given task. Although the double dissociation approach should prevent this from happening, the fact that the brain is organized in a quite decentralised fashion finally led to the insight that most cognitive functions are organised in a "functional network" with identifiable sub modules. Thus, if for example as a result of a brain injury, one brain region is damaged and the cognitive impairment becomes smaller and smaller over time, this will be because another brain area will fulfil the required task and thus take over the role of the damaged module (a very impressing case of plasticity is described in Muckli et al. 2005).

Some achievement is also made by manipulating the brain at the level of brain areas. This can either be done by hindering a certain area in participating for a certain task by "transcranial magnetic stimulation" (TMS) or even by directly inducing a pulsed current with the help of "brain pacemakers". This so called "deep brain stimulation" (DBS) method is used in the treatment of Parkinson and more recently even in the treatment of chronic depression (Mayberg et al. 2005). Accordingly, it is possible to change a cognitive state like a patient's mood by a relatively unspecific[44] stimulation of a certain brain area (in this case, Brodmann area Cg 25).

Still, neuroscientists become more and more aware of the fact that the spatial resolution of brain areas is too coarse to answer the most interesting question (Elger et al. 2004): What makes a brain state a brain state with a certain content and how does the brain code information? Note that by this question, the taxonomy criteria is already included: content. By this, neuroscientists do not only ask for the phenomenal content of conscious

[44] Current and frequency play an important role but this method does of course not intervene on the level of single neurons.

experiences –about which a huge discussion can also be held (Metzinger 2000) – but for how neural nets code information in general.

There are several possible answers to this question. The easiest and most intuitive one seems the proposal of a "mapping" which means nothing else than a neuron coding information due to its location and projections. An example would be an anecdotic called "grandmother neuron" which is supposed to carry the information "grandmother" (e.g. there is) just by firing of this very neuron. Along these lines one would be able to correlate the presence of a certain mental content with the firing (or way of firing when one regards that all neurons spontaneously fire) of a certain neuron. Though this picture was often used to illustrate how neural networks *do not* code information, recent research has been able to identify such highly specific neurons (Quiroga et al. 2005). We also would expect that selective stimulation of such a neuron would evoke a thought about ones grandmother. The attractiveness of this picture about neuronal coding is based on its intelligibility: We would be able to directly read the thought of the bearer of the brain that we investigate. What is possible at a very low level of visual processing, is that one can find and thus interpret a "copy" of the activation of the retina in the primary visual cortex, would be possible on a higher cognitive level as well. As soon as one had the mapping available, one would just have to look at which neurons are active and the mental content would be known. Thus, from the perfect information about someone's brain, we could not only directly "read" if he sees a vertical or a horizontal bar, but also if he thinks about a special person.

The problem with this account for the purpose of psychophysical identities is – besides the fact that such highly specifically responding neurons are rather the exception than the rule – the same as with the C-Fibre example. Though they might *represent* a certain information *for* another cognitive system, their activity can not be identified with a mental content: Extract that neuron, induce a current and without an "interpreter" that "*takes*" the activity of that neuron as being the information, there is no such feature of that neuron.

There is a lesson in this: From the possibility to manipulate such selective neurons, as well as from the possibility to manipulate highly selective brain areas, we learn that activity of such a neuron or activity in such a brain area is a necessary condition for the bearer of the corresponding brain to be in a mental state with the content that is coded by this neuron or area. That this is a necessary condition is due to the embedding of the neuron or area in a bigger functional network. Unfortunately, it seems to be far from sufficient. This problem also applies for almost all other candidates for being the neural type in a psychophysical type identity as well.

But what about a candidate which most scientist agree about that it has something to do with making a neural activation pattern a pattern of a certain type – the formation of a neural ensemble? The idea goes back to Donald Hebb (1949) and due to the increasing insight on the distributional coding of neural nets, it becomes apparent that a certain information can be coded by a network of neurons which are distributed across the brain and are activated in a certain pattern. Again, we are asking for what state we would regard to fulfil *all* necessary conditions so that we could agree that this state realises e.g. a belief, a desire or a sensation. The question about a system's sufficient conditions to exhibit a mental content is not new and obviously we are still unable to give an exact answer. But from understanding of how the brain codes information with the additional knowledge of which functional modules exist, we can generate a good hypothesize of how the answer will look like.

On the one hand we have to be able to decode the neural activation patterns of the cell ensembles involved in order to have knowledge of the exact content and on the other hand, we have to know what happens with these representations of information and which functional brain modules are influenced by or involved in these activation patterns. If we know all this, we would also know which features the system has to exemplify and in which "activity pattern" it has to be to instantiate mental states. Thus, the sufficient conditions for being in a certain psychological state are extremely complex, still, there is definitely good reason to

believe that we can get a grasp on the sufficient conditions for a system to be in a specific mental state.

The "litmus test" for this claim would of course be a situation where perfect information about the exact activity patterns of brains is available and we could exactly predict mental contents. For very low level content, like colour perception, this can already be done (Haynes and Rees 2005). For more complex content which involves behavioural dispositions and a greater number of subjective associations, there are more necessary features of the activity pattern (i.e. it might have to span over more brain areas, be synchronised in a certain way…) and the network in which the pattern takes place. In a line with this, it has been shown that higher cognitive functions lead to widely spread synchronous firing of neurons compared to the relatively restricted synchrony found in low level stimulus processing (Ward 2003; also see Hagoort et al. 2004). This is probably not even surprising. When we believe it to be essential for a desire to have a behavioural consequence, there has to be a way by which the pattern, which realises that desire, can influence the motor areas. At the end, to be in a certain cognitive state will be identifiable with being in a certain neural state which means, that the neurons have to exhibit a certain activation pattern where for the neural dynamics, such things as frequency and synchrony will play an essential role.[45]

By this, we have answered the above question for neural tokens – at least in a very general way. What is left to answer is if these tokens naturally group in a way so that one can speak of a type. This is exactly what Fodor (1974) doubted. As a materialist, one could say: "Of course they do, because instantiations of pain instantiate pain and instantiations of tickle instantiate tickle, pain tokens have to group on some abstract level as well as tickle tokens." But this would of course beg the question since this is exactly what has to be shown. In the imaginary case of the availability of the perfect information about someone's brain activity,

[45] Speaking about states makes it difficult to consider "temporal coding", this might indeed be a reason to favour speaking about the identity of processes.

we could devise an experiment that should overcome this problem. In order to commence in a simple manner, we will not think of all mental states but just of states as being in pain and feeling a tickle. What we would do then is to inflict pain and tickle to the participants of our experiment. With the help of our perfect brain scanner, we always save the exact activation pattern when we inflict the mental states to them, either by twitching them or by tickling them, meanwhile we ask them if their introspection reveals if they are in the mental state they are supposed to be in. What we get then is a very huge collection of data. Each neural token – the record we made while our participant was in one of the two states – can be regarded as a very high dimensional vector that expresses the exact state of each neuron. All vectors together span an "input space" – the space of every possible activity pattern. What we want to know now is if the tokens that are equal with the mental tokens which group into two categories (pain and tickle) are "separable" in a way that two corresponding categories arise. For such a categorisation or clustering task, artificial neural networks (especially "Self Organising Maps" (SOMs) Zell 1994)[46] are perfectly suited. Formulated in another way, the question is whether a network can be found that is trainable to such a clustering without exception.[47]

Could Fodor argue now, that the input space of the neurological tokens is not separable in the way that all neurological tokens that are identified with pain group together and all neuro-tokens that are identified with tickle are grouped together? After all, if a human

[46] One problem has to be admitted that holds for most architectures of artificial networks: With respect to aiming at a certain outcome, it does often play an important role how the network is construed. Consequently, one has to take care that one does not put something into the model that should be shown. To lay down a network that separates an input space that is not separable is still not possible.

[47] Though I tried to incorporate actual considerations of neuroscience, the separability argument of micro level tokens, as one might call it, can of course also be formulated with other hypothesis about the nature of the reduction base of mental phenomena as well. Even reference to artificial neural networks is not the only way to lay out a version of my argument. Other criteria for the similarities and differences of tokens might do the job as well.

reliably reports that he is in pain when there is one of the activity patterns in is brain that we recorded during pain situations, this means that the input space is separable!

The situation in which there will be a discrepancy between the clustering of the tokens on the neural level and the categories in which their mental counterparts are grouped will be important for the next section. What could happen is that the clustering can actually be finer grained in the way that the neurological counterparts to pain tokens cluster in more than one distinct type (which still do share some commonalities). Maybe it would even be possible that the tokens recorded in our pain/tickle experiment group in three ore more quite different clusters. Under these conditions it would be interesting to which kinds of mental tokens the tokens of the third neural type would project. I see two possibilities here: either the participant reliably reports that she is in one of the two states (e.g. pain) when she is in neural state that is really quite distinct from the one also identified as a pain state or she reports by chance that he is in one of the two states i.e. in another trail he might report otherwise. In the first case, there can again be two different explanations. Either we would impute that the participant suffers from a linguistic confusion – he uses the word pain ambiguously for two different types of states. The second possibility would be that what we all mean by the mental type pain is indeed a disjunction of two different states. In the second case, we would indeed assume that the participant just does not have a word or category for the third state so that – in a case of forced decision – he will use the two available words haphazardly.

VI.3 Considerations about Psychological Types

We have finally come to the point where I can confer my example of the taxonomy of life-forms to the mind-body domain. Although, I have pointed out that it does not really play a role if one talks about scientific psychology or folk psychology in some cases, the real challenge is to deal with folk psychology. There are two reasons for this. First, the mind-body-problem arises exactly from the difference in perspective onto the investigated phenomena. While the first person perspective is described as “privileged” or “non-

observational", science is done from the third person perspective and is observational. Scientific psychology is no exception here for the most part. It is largely accepted that introspection delivers only very unreliable data and thus scientific psychology often refers to categories like processing steps measured by reaction time that do not really play a role in folk psychology. Second, it is palpable that my argument regarding the scientific taxonomy is not in conflict with scientific psychology. The refinement of scientific types is usual practice in scientific psychology. Alongside the example of mental illnesses, the case of the memory system is a good example: At some time it became apparent that the distinction between long term and short term memory enables explanation for phenomena that could not be done before that distinction (Eysenck and Keane 2000, pp. 151ff). Then again, the concept of short term memory was replaced by the concept of a working memory that could be described in even smaller modules which enabled that explanation of even more phenomena.

Is folk psychology really so different then? After all, scientific psychology and even neurology increasingly influence the use of psychological terms in every day language. To differentiate between two different memory modules is already part of common sense and the same assimilation of scientific knowledge can also be observed in other examples of folk understanding of the mental.

Still, there seems to be a difference between what philosophers think that mental types are and respectively how they come to be on the one side and the slippery types of science on the other side. The reason for this was already mentioned in the original Smart article (1959, p. 152) as one objection against the identity thesis:

> Sensations are private, brain processes are public. If I sincerely say, "I see a yellowish-orange after-image" and I am not making a verbal mistake, then I cannot be wrong.

This conviction seems still to be shared by many philosophers. What should be kept in mind is that it was a type-type version of the identity theory that was under discussion here. Consequently we have to attest that there was the opinion that one cannot be wrong about the

own mental types – in some way, they were thought to be dogmatically given. Smart only formulated the objection for sensations but for desires and beliefs there seems to be the same intuition as well: the access to one's own mental states is perceived as a direct access to metaphysical facts. This is exactly the point were I regard the insignificance claims about the token identity thesis as mislead. First, it reminds us of the distinction between types and tokens and second it gives way for acknowledging what scientific types really are: abstractions from concrete instances and thus denoting bundles of tokens. The interesting question is how folk psychological types are generated. And my answer is that they are generated from the tokens one encountered so far. In section V.1 I already pointed out that types are not directly 'given' but arise from confining the tokens. So there are instantly types. But this does by no means indicate that there are natural kind types. And only the identity of natural kind types is really interesting for reductive accounts since non natural kind types are at best disjunctions of natural kind types and in his "Special Sciences" article (Fodor 1974), Fodor was right concerning the difficulties of type disjunctions. So we do know how mental types come into existence but it is more interesting how natural kind types can be grasped. Fodor's claim was that special sciences do talk about natural kinds but that they do not have to correspond to a natural kind of a lower level science. Of course the whole debate applies only in a situation where there are no errors in scientific knowledge anymore – in a situation were an exception to a scientific law is not anymore due to mistakes in the law. And under this situation, I definitely doubt Fodor's claim. To draw the lines between the tokens in the manner that the resulting type is a natural kind is exactly to draw them in a way that the corresponded tokens on the lower level science group into corresponded types! Thus, to generate a natural kind mental type, one has to take care of all the natural differences that can be found and this includes taking the micro level difference into account.

VII Implications for the Reliance in Introspection

This is the last chapter before I will come to a conclusion. Though my argument is brought forward and I hope to have illustrated the way in which it applies to the taxonomy of neurology and psychology, I want to say something more about how my picture of scientific types stand in contrast to the claim that one can not be mistaken about the own mental states. Well, if we talk about mental tokens, this might even be true but the claim seems to be a tautology then. It would mean nothing more that the actual mental state has the properties it has. Of course does a sensation feel like it does, has a believe the content that is has and does a desire motivate in the way it does. And that is why the claim has this extreme appeal of plausibility. The identification of the tokens with physical tokens does not make the slightest difference here. But the claim was actually not about tokens, it was about types. And as my previous remarks about “primal types” and proto taxa based on intuition should have made clear: the opinion about which natural kind types there are *can* be mistaken. The claim about the impossibility of mistakes to self ascriptions explicitly excluded verbal mistakes but one probably has to ask if this makes sense in the case of types. As abstractions, types are quasi pre-lingual entities and the exclusion of verbal mistakes thus seems to exclude (folk-) scientific mistakes. Thus, examples that come to mind do always leave the question if it is not a verbal mistake. What about a child who swears that it is not tired while the opposite is obviously true? The question if it really believes that it is not tired and thus does not make a verbal mistake is not that clear. Someone coming to the doctor reporting pain where as the doctor diagnoses after some questions that this is actually not pain but the patient is probably depressive, might have honestly classified the sensation during depressive periods as painful. Only that the majority of society started at some point to differentiate between pain and feeling abject. And what about someone who is thought in music and is finally able to hear smaller and smaller nuances of tone differences he did not her before his training? When he reports about the sensation of hearing a short melody, he can probably differentiate as well

between the sensations that he has while hearing two different octaves that he described as the same before.

As a consequence to these considerations, the above argument portrayed by Smart can either be regarded as a tautology or as simply wrong. Either a yellowish-orange after-image is a type whose boarders can change or this is excluded as a verbal mistake. As the example of a colour sensation type is exemplarily in the argument, the flaw of the argument is the same for all mental types.

VIII Conclusion: Scientific Taxonomy results in Type Identity if Token Identity holds

The aim of this thesis was to demonstrate that the death announcement of the identity theory of mind and body is a consequence of a misconception of what types are and how types come about in the sciences. We have learned about the three most influential arguments against psychophysical type identities and now, after I sketched a picture about what types really are – namely post hoc structures over bundles of tokens – it is time to check if the three arguments still have significance.

When it becomes acknowledged that it is partially a question of definition which tokens are subsumed under a type it also becomes apparent that the multiple realizability argument holds only if types on the to be identified with levels are defined in the way that there will be a mismatch. As I have shown, there is good reason that a “natural grouping” on the two levels will generally lead to types of a similar degree of abstraction. But maybe even more important is what would happen if there are still mismatches in the way that the tokens on the neural level make it possible to differentiate between types that on the mental level are all classified as one type: then we would have good reason to differentiate between the mental tokens that project to different subtypes on the neural level too. In other words, if there is reason to believe that on one level the natural kinds are finer grained, then there will be a way of discovering slight differences of subtypes on the other level as well. Along this line, it can

be beautifully explained what the mistake is when using the multiple realizability argument against psychophysical identities and thus the argument is misleading.

What about the necessity argument then? Well, we can now say which of Kripke's premises were wrong. In his time, the identification with things like C-Fibres really made the impression of a contingent fact. In fact, if we generate the types as I have described, the identity between them is indeed a necessary one. First, the identification of a mental token with a neurological token as I have described it, is indeed not contingent. If the neural architecture and the activation pattern that are present in it, hold the sufficient conditions, to be in this state really is to also be in a state that exhibits the mental properties. To claim the opposite is as to claim that an artefact that has the same properties and is in the same state as a car with a combustion engine that moves due to explosions in its engine moves only contingently forward when there is an explosion in one of the cylinders.[48] Second, the practise of scientific taxonomy secures that a higher level type will be "adjusted" when on the ground of knowledge about the micro level there is reason to change the contours of the higher level type. Thus, if the micro level is sufficiently investigated, the macro level types will be a one to one projection to the micro level.

I have to admit that the explanatory gap argument can not be disposed so easily. In fact, the argument is only touched at the edges. But even here, I am not that pessimistic. The thing is that most advocates of the argument do not deny that is might be possible to completely link neural activity to behaviour. They only deny that the qualitative properties of mental states can be explained by neurology. As already mentioned, I do have some trouble separating behaviour from phenomenal content in this way. Furthermore, I can well imagine that at some time the neurologist will claim that they understand why being in a specific neural condition has to feel as it does. The supporters of the argument do also have to show that they do not just "play possum". Eventually it is easy to say: "I don't get it." Without the

[48] In other words it is as necessary as the natural laws are.

necessary mathematical knowledge and practice, the claim of a highly skilled mathematician that two formulas can be converted into each other can also not be comprehend by myself. And as nobody would deny, a realistic account of the neural activity of the brain will be so complex that the *impression* of understanding might indeed not arise. Probably the situation might end up like in quantum physics where for most physicists it is true that they can describe the processes but only few will claim that they have "understood" them. Clearness decreases dramatically as the mathematically description increases in dimensions.

In this thesis I have tried to summarize the most important arguments for why type identity claims are too ambitious. On one side, I tried to show that these arguments were wrong about what types are and how they come to be. One the other side I discussed in section IV.3. why token identity was thought not to deliver the crucial expressiveness to be of much relevance in giving interesting answers to the mind-body problem. The deciding point here is that this might be true if the token identity thesis is examined in isolation. This token identity thesis though, does not stand in isolation. It is the framework of science in which explanations take place and scientific practice ensures that materialism, as well reductionism and type physicalism hold when token identity holds. This is done by refining types until they match natural kinds, which again means that types are reducible.

At the end, I like to conclude: There's life in the old dog – also known as the identity theory of mind and body –yet!

IX References

Abbott, Barbara. 1997. A Note on the Nature of "Water". *Mind* 106 (422):311.

Acton, G. Scott, and Jason J. Zodda. 2005. Classification of Psychopathology: Goals and Methods in an Empirical Approach. *Theory Psychology* 15 (3):373-399.

Baker, Lynne Rudder. 1987. *Saving belief: a critique of physicalism*. Princeton, N.J.: Princeton University Press.

Batterman, R. W. 2000. Multiple realizability and universality. *British Journal of Philosophy of Science* 51 (1):115-145.

Bechtel, William, and Robert N. McCauley. 1999. Heuristic Identity Theory (or Back to the Future): The Mind-Body Problem Against the Background of Research Strategies in Cognitive Neuroscience. In *Proceedings of the 21st Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bechtel, William, and Jennifer Mundale. 1999. Multiple Realizability Revisited: Linking Cognitive and Neural States. *Philosophy of Science* 66 (2):175.

Beckermann, Ansgar. 2001. *Analytische Einführung in die Philosophie des Geistes*. 2., überarbeitete Aufl. ed. Berlin; New York: W. de Gruyter.

Block, Ned. J. 1995. On a Confusion about a Function of Consciousness. *Behavioral and Brain Sciences* 18 (2):227-247.

Carrier, Martin, and Jürgen Mittelstraß. 1995. *Mind, brain, behavior*. Translated by S. Lindberg. Rev. and expanded Engl. ed. ed. Berlin; New York: de Gruyter. Original edition, Geist, Gehirn, Verhalten.

Chalmers, David John. 1996. *The conscious mind: in search of a fundamental theory, Philosophy of mind series.* New York: Oxford University Press.

Cheyne, Colin. 1993. Reduction, Elimination, and Firewalking. *Philosophy of Science* 60 (2):349.

Churchland, Patricia Smith. 2002. *Brain-wise: studies in neurophilosophy*. Cambridge, Mass.: MIT Press.

Churchland, Paul M. 1989. Folk Psychology and the Explanation of Human Behavior. *Philosophical Perspectives* 3:225.

———. 1992. *A Neurocomputational Perspective*. Cambridge, Mass.: MIT Press.

Cruse, Holk. 2004. Ich bin mein Gehirn. Nichts spricht gegen den materialistischen Monismus. In *Hirnforschung und Willensfreiheit*, edited by C. Geyer. Frankfurt am Main: Suhrkamp.

Davidson, Donald. 1970. Mental Events. In *Experience and Theory*, edited by L. Foster and J. W. Swanson.

———. 1973. The Material Mind. In *Proceedings of the fourth international congress for logic, methodology, and philosophy of science*, edited by P. S. e. al.

———. 1974. Psychology as Philosophy. In *Philosophy of Psychology*, edited by S. C. Brown. London/Basingstoke: Macmillan.

———. 1980. *Essays on actions and events*. Oxford: Oxford University Press.

Descartes, Renâe. 1986. *Meditations on first philosophy: with selections from the objections and replies*. Cambridge [Cambridgeshire]; New York: Cambridge University Press.

Dobbs, David. 2005. Fact or Phrenology? The growing controversy over fMRI scans is forcing us to confront whether brain equals mind. *Scientific American Mind* 16 (1):24-31.

Elger, Christian E., Angela D. Friederici, Christof Koch, Heiko Luhmann, Christoph von der Malsburg, Randolf Menzel, Hannah Monyer, Frank Rösler, Gerhard Roth, Henning Scheich, and Wolf Singer. 2004. Das Manifest; Elf führende Neurowissenschaftler über Gegenwart und Zukunft der Hirnforschung. *Gehirn & Geist* 6.

Eysenck, Michael W., and Mark T. Keane. 2000. *Cognitive psychology: a student's handbook*. 4th ed. Hove, East Sussex, UK: Psychology Press.

Feigl, Herbert. 1967. *The "mental" and the "physical"; the essay and a postscript*. Minneapolis: University of Minnesota Press.

Fodor, Jerry A. 1968. *Psychological explanation; an introduction to the philosophy of psychology*. New York: Random House.

———. 1974. Special Sciences, or the Disunity of Sciences as a Working Hypothesis. *Synthese* 28:97-115.

———. 1981. *Representations: philosophical essays on the foundations of cognitive science*. 1st MIT Press ed. Cambridge, Mass.: MIT Press.

———. 1987. *Psychosemantics: the problem of meaning in the philosophy of mind, Explorations in cognitive science*. Cambridge, Mass.: MIT Press.

———. 1991. You can Fool Some of The People All of The Time, Everything Else Being Equal; Hedged Laws and Psychological Explanations. *Mind* 100 (397):19-33.

Foster, John. 1994. The Token-identety Thesis. In *The mind-body problem: a guide to the current debate*, edited by R. Warner and T. Szubka. Cambridge, Mass.: Blackwell.

Hagoort, Peter, Lea Hald, Marcel Bastiaansen, and Karl Magnus Petersson. 2004. Integration of Word Meaning and World Knowledge in Language Comprehension. *Science* 304:438-441.

Haslam, Nick. 2002. Kinds of Kinds: A Conceptual Taxonomy of Psychiatric Categories. *Philosophy, Psychiatry, & Psychology* 9 (3):203-217.

Haynes, John-Dylan, and Geraint Rees. 2005. Predicting the Stream of Consciousness from Activity in Human Visual Cortex. *Current Biology* 15:1301–1307.

Hebb, D. O. 1949. *The organization of behavior; a neuropsychological theory, Wiley book in clinical psychology*. New York: Wiley.

Hunn, Eugene. 1975. A Measure of the Degree of Correspondence of Folk to Scientific Biological Classification. *American Ethnologist* 2 (2):309.

Jackson, Frank. 1986. What Mary Didn't Know. *The Journal of Philosophy* 83 (5):291.

Kandel, Eric R., James H. Schwartz, and Thomas M. Jessell. 2000. *Principles of neural science*. 4th ed. New York: McGraw-Hill Health Professions Division.

Kim, Jaegwon. 1966. On the Psycho-Physical Identity Theory. *American Philosophical Quarterly* 3:225-235.

———. 1989. The Myth of Nonreductive Materialism. *Proceedings and Addresses of the American Philosophical Association* 63 (3):31-47.

———. 1992. Multiple Realization and the Metaphysics of Reduction. *Philosophy and Phenomenological Research* 52 (1):1.

———. 1994. Supervenience. In *A Companion to the Philosophy of Mind*, edited by S. Guttenplan. Oxford: Blackwell.

———. 2005. *Physicalism, or something near enough, Princeton monographs in philosophy*. Princeton, N.J.: Princeton University Press.

Kripke, Saul A. 1971. Identity and Necessity. In *Identity and individuation*, edited by M. K. Munitz. New York: New York University Press.

———. 1980. *Naming and necessity*. Cambridge, Mass.: Harvard University Press.

Lenzen, Wolfgang. 1998. Zombies, Zimbos und das »schwierige Problem« des Bewußtseins. In *Bewubtsein und Repräsentation*, edited by H.-D. Heckmann and F. Esken. Paderborn [u.a.]: Schöningh.

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

———. 2002. Gedanken über Qualia. In *Phanomenales Bewusstsein-Ruckkehr Zur Identitatstheorie?* edited by M. Pauen and A. Stephan. Paderborn: mentis.

Lewis, David. 1994. Reduction of Mind. In *A Companion to the philosophy of mind*, edited by S. D. Guttenplan. Oxford, OX, UK; Cambridge, Mass., USA: Blackell Reference.

Mayberg, Helen S., Andres M. Lozano, Valerie Voon, Heather E. McNeely, David Seminowicz, Clement Hamani, Jason M. Schwalb, and Sidney H. Kennedy. 2005. Deep Brain Stimulation for Treatment-Resistant Depression. *Neuron* 45:651–660.

Metzinger, Thomas. 2000. *Neural correlates of consciousness: empirical and conceptual questions*. Cambridge, Mass.: MIT Press.

Muckli, L.F., M.J. Naumer, R. Sireteanu, and W. Singer. 2005. Bilateral visual field representation in V1 of a patient born with only one hemisphere. Paper read at Society for Neuroscience Meeting, at Washington DC.

Pauen, Michael. 2000. Identität und multiple Realisierung: ein prinzipieller Gegensatz? Paper read at GAP.4 Argument und Analyse, at Bielefeld.

Place, U.T. 1956. Is consciousness a brain process? *British Journal of Psychology* 47:44-50.

Putnam, Hilary. 1960. Minds and Machines. In *Dimensions of mind; a symposium*, edited by S. Hook. New York: New York University Press.

———. 1967a. The Mental Life of Some Machines. In *Intentionality, minds, and perception; discussions on contemporary philosophy, a symposium*, edited by H.-N. Castañeda. Detroit: Wayne State University Press.

———. 1967b. The Nature of Mental States. In *Mind, language, and reality*. Cambridge Eng.; New York: Cambridge University Press.

———. 1967c. Psychological Predicates. In *Art, Mind and Religion*, edited by H. W. Capitan and D. D. Merrill. Pittsburgh: Pittsburgh University Press.

Quine, W. V. 1969. Natural Kinds. In *Ontological relativity, and other essays*, edited by W. V. Quine. New York: Columbia University Press.

Quiroga, R. Quian, L. Reddy, G. Kreiman, C. Koch, and I. Fried. 2005. Invariant visual representation by single neurons in the human brain. *Nature* 435:1102 -1107.

Ronald, P. Endicott. 1993. Species-specific properties and more narrow reductive strategies. *Erkenntnis* V38 (3):303.

Root, Michael. 2000. How We Divide the World. *Philosophy of Science* 67:S628.

Schwartz, Justin. 1991. Reduction, Elimination, and the Mental. *Philosophy of Science* 58 (2):203.

Smart, J. J. C. 1959. Sensations and Brain Processes. *The Philosophical Review* 68 (2):141.

Stephan, Achim. 1999. *Emergenz. Von der Unvorhersagbarkeit zur Selbstorganisation*. Paderborn: mentis.

van Gelder, T. J. 1998. The roles of philosophy in cognitive science. *Philosophical Psychology* 11:117-136.

Ward, Lawrence M. 2003. Synchronous neural oscillations and cognitive processes. *TRENDS in Cognitive Sciences* 7 (12):553-559.

Wittgenstein, Ludwig, and G. E. M. Anscombe. 2001. *Philosophical investigations: the German text, with a revised English translation*. 3rd ed. Oxford; Malden, Mass.: Blackwell.

Zachar, Peter. 2000. Psychiatric Disorders Are Not Natural Kinds. *Philosophy, Psychiatry, & Psychology* 7 (3):167-182.

Zell, Andreas. 1994. *Simulation neuronaler Netze*. 1. Aufl. ed. Bonn [u.a.]: Addison-Wesley.