Psycho-neural reduction through functional sub-types

Patrice Soom, Christian Sachse & Michael Esfeld[1]
University of Lausanne, Department of Philosophy
(forthcoming in *Journal of Consciousness Studies*)

Abstract

The paper argues that a functional reduction of ordinary psychology to neuropsychology is possible by means of constructing fine-grained functional, mental sub-types that are coextensive with neuropsychological types. We establish this claim by means of considering as examples the cases of the disconnection syndrome and schizophrenia. We point out that the result is a conservative reduction, vindicating the scientific quality of the mental types of ordinary psychology by systematically linking them with neuroscience. That procedure of conservative reduction by means of functional sub-types is in principle repeatable down to molecular neuroscience.

1. Functional reduction

There are good arguments for taking the mind to be identical with the brain. In other words, any mental property token as described by psychology is identical with some configuration of neurobiological property tokens that can in principle be identified and described by a neuroscientific theory. It is common ground that (1) the mental strongly supervenes on the physical, that (2) every physical effect has a complete physical cause insofar as it has a cause at all and that (3) mental property tokens are causally efficacious. If one takes mental tokens not to be identical with configurations of neurobiological tokens, which are a sort of physical tokens, then one runs into what is known as the causal exclusion problem: assuming that a mental property token m_1 causes another mental property token m_2 (3), m_1 can bring about this mental change only by bringing about a physical change as well (1), that is, by having a physical effect p_2 that is sufficient as a supervenience basis for m_2 (given suitable background conditions). However, given (2), there always is a complete physical cause p_1 for p_2:

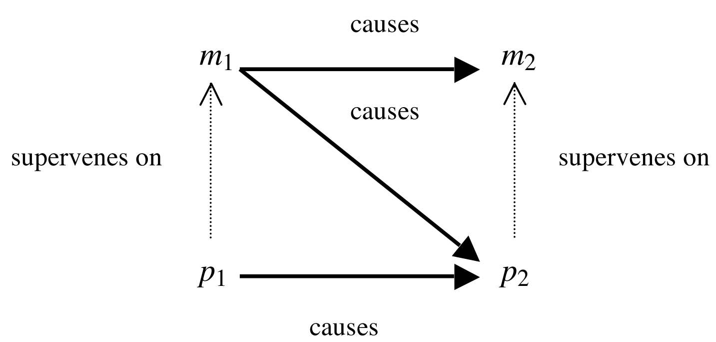

Figure 1: The causal exclusion problem

1 Patrice Soom gratefully acknowledges support from the Swiss National Science Foundation (SNF), grant nr. PDFM1-118603/1.

Facing this situation, it is possible to maintain that mental causes overdetermine all the effects that they cause: for any effect of a mental cause *m*, there always is a physical cause *p* that is sufficient to ensure the existence of the effect in question. However, such a situation of overdetermination is indistinguishable from a situation in which the mental is epiphenomenal (that is, does not cause anything). Overdetermination therefore collapses into epiphenomenalism: there always are sufficient physical causes, and the physical laws prevail over the psychological or psychophysical ones, since they are strict and universal laws rather than *ceteris paribus* laws that admit of exceptions. The causal exclusion problem therefore is a strong argument for taking all mental property tokens to be identical with physical property tokens (Kim, 1998). If there is mind-brain identity, the problem of mental causation vanishes: some configurations of neurobiological property tokens are mental property tokens so that mental property tokens are causally efficacious because they are themselves something neurobiological.

However, taking the causal argument to show *that* there is mind-brain identity does not answer the question *how* it is possible in concrete terms for the features that characterize something qua mental property token to be identical with a certain configuration of neurobiological tokens. The most promising answer to that question is functional reduction. The general account of functional reduction, also known as Ramsey-Lewis reduction, runs as follow. The first step consists in the functionalization of mental properties by defining them in terms of a characteristic causal role. Following Lewis (1994), common sense psychology conceives mental properties in terms of their characteristic effects (which can be mental as well as physical). The causal role in question is realized by certain configurations of neurobiological property tokens. The discovery of these configurations that make true the functional description of step (1) constitutes the second step. The description of these neurobiological configurations is no longer couched in a mental vocabulary, but in a neuroscientific vocabulary, and, in the last resort, in a physical vocabulary (say, of molecular neuroscience). Finally, it is in principle possible to explain in neuroscientific terms how the causal role is exercised that the mental description brings out in functional terms.

Functional reduction thus shows how in each individual case a mental property token is identical with a configuration of neurobiological property tokens that brings about the effects due to which the token in question comes under a certain mental description. Functional reduction hence gives rise to lawlike, psychophysical generalizations: any neurobiological property token of the same type as the considered one, any N_1, makes true an M, that is, a mental description of the same type as the considered one:

(1) $\forall x\ (N_1 x \rightarrow Mx)$.

The symbol "∀" designs the universal quantification over a domain of individuals. Proposition (1) can be read in English as "For any property token x, if x falls under the neurobiological description N_1, then x falls under the psychological description M.

However, functional reduction faces a major problem: not any property token that makes true a mental description of the type M makes that description true because it makes true a neurobiological description of the type N_1. The reason is that the mechanism to bring about the effects that define M does not have to be of the type N_1, but it can also be of the type N_2. This is what is known as multiple realization: configurations of property tokens of different neurobiological types N_1, N_2, N_3, etc. can all make true a mental description of one and the same type M because there are different mechanisms that all constitute minimal sufficient conditions to bring about the effects that define M. Consequently, there are also lawlike psychophysical generalizations of the type:

(2) $\forall x\ (N_2x \rightarrow Mx)$.

Hence, multiple realization excludes bi-conditionals between neurobiological and mental types of the following form:

(3) $\forall x\ (N_1x \Leftrightarrow Mx)$.

Functional reduction, as it stands, does therefore not amount to a theory reduction of psychology to neurobiology, and the identity of any mental property token with something neurobiological is stuck in limbo, for it is not supported by a type identity in the sense of a coextension between mental property types and neurobiological property types.

Reacting to this situation by taking multiple realization to show that mental property tokens are after all not identical with configurations of neurobiological property tokens is no way out: they would then be epiphenomenal, as the problem of mental causation teaches us. The strategy of Lewis and Kim is to search for species-specific realizers of a given mental property type: the idea is that all the configurations of neurobiological property tokens that make true a description of the mental type M in a given species (or sub-species or still smaller group) are tokens of the neurobiological type N_1. In that manner, psychology would be reduced to the neurobiological theory N_1 in species S_1, to N_2 in species S_2, etc. (Lewis, 1980; Kim, 1998: 93-95, 2005: 25). However, the unity of psychology gets lost in this strategy: there is no place for a psychology that seeks for generalizations that are valid across different species; psychology is in each case replaced with a neurobiological theory that takes its place for a certain domain, that is, a certain species, sub-species or even smaller group (cf. the so-called new wave reductionism of Bickle, 1998). The general concepts or types of psychology do not have any scientific value (e.g. Kim, 1999: 17-18). We thus face a dilemma of either epiphenomenalism for mental property tokens or eliminativism of the scientific value of psychological types.

A new strategy for developing functional reduction into a fully-fledged reduction despite multiple realization has been developed recently, seeking to avoid that dilemma by means of a conservative theory reduction (Esfeld and Sachse, 2007). That strategy is motivated by the observation that multiple realization arises from the fact that the mental, functional types of psychology are coarse-grained, whereas neurobiological types are much more fine-grained (Bechtel and Mundale, 1999). This is why the mental, functional types of psychology have a larger extension than the neurobiological ones. Against this background, the central idea of the new strategy is to define in the vocabulary of psychology more precise functional, mental sub-types that are in the end as fine-grained as the neurobiological types. The argument can be summed up in the following manner (Esfeld and Sachse, 2007: 5):

1) Assume that two configurations of neurobiological property tokens, n_1 and n_2, fall under the neurobiological descriptions N_1 and N_2 and both make true the mental, psychological description M. There is in this case a systematic difference in the way in which n_1 and n_2 bring about the effects that define M. In other words, the types N_1 and N_2 capture two different mechanisms of producing the effects that define M.

2) Any such difference in mechanisms leads to a systematic difference in the production of functional side-effects that are linked to the production of the effects characterizing M. For any of these differences, there are physically possible conditions under which those side-effects manifest themselves in a functional salient way, that is, in a way that is salient for psychology. Hence, it is possible to grasp those differences in a functional manner by making the functional, mental types of psychology more precise in their own vocabulary.
3) On the basis of those side-effects and their manifestation under critical conditions, functionally defined sub-types M_1, M_2, M_3, etc. of M can be constructed. All those sub-types include the definition of M but are individuated by taking in a functional way the side-effects peculiar to each neurobiological type N_1, N_2, N_3, etc. into account. For any functional, mental description M and any neurobiological description N_1 whose referents make true M, it is in principle possible to conceive a functional sub-type M_1 (of M) that is co-extensional with N_1.

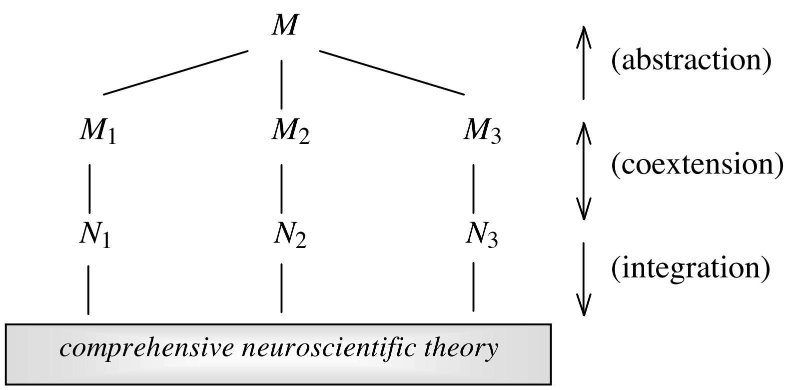

Figure 2: The reduction of psychology to neurobiology via functional sub-types

Note that this procedure is repeatable. We can in principle construct functional sub-type M_{1a} of functional sub-type M_1 of M until our final sub-types are no longer multiply realizable, that is, are coextensive with neurobiological types. This improvement of functional reduction leads to the following two results: (1) It provides for (sub-)type-identities in the form of biconditionals, thereby making the reduction of psychology to neuroscience possible. (2) It yields a conservative reduction, establishing the scientific quality of the general types of psychology (and the lawlike generalizations couched in terms of them); for these are contained in the sub-types that are coextensive with neurobiological types and simply abstracted from them. Hence, psychology is vindicated as bringing out salient similarities across different species that neurobiology does not have the conceptual means to express, and psychology is systematically linked with neurobiology by means of a conservative, functional reduction via functional sub-types.

That strategy of functional reduction has been applied to biology hitherto (Sachse, 2007). The goal of this paper is to extend it to the relationship between common sense psychology and neuropsychology. In sections 2 and 3, we will introduce a very simple example in order to show how this model can be applied to the relation between functionalized folk psychology and neuropsychology. Section 4 will then be concerned with the implementation of the sub-types strategy of reduction in a case of more complex neuropsychopathology.

2. *Disconnection syndrome: split-brain patients*

The concept "inter-hemispheric disconnection syndrome" picks out a set of behavioural abnormalities arising, from the neurological point of view, from injuries or from a total lack of corpus callosum. The corpus callosum is a large medullary strip linking both hemispheres of mammalians brain allowing information exchange between the two hemispheres. It is composed of circa 200 to 800 millions commissural fibres that can be tied up in three main classes (Kolb and Whishaw, 2003: 428-430). Most of them are topographic in the sense that they connect nervous areas to their respective contralateral counterpart. A second group of connections goes to areas to which the homotopic area on the contralateral side projects and, finally, a last group of connections has diffuse terminal distributions.

Damages to the corpus callosum can occur accidentally consequent upon congenital malformations or surgical ablation. Corpus callosum's natural agenesis is one of the most common brain malformations observed in humans with a prevalence of 3-7 per 1000 individuals (Bedeschi et al., 2006). The surgical removal of the corpus callosum, the therapeutic commissurotomy, has been reintroduced in the early sixties of the last century by the surgeons Philip Vogel and Joseph Bogen as an elective treatment for severe cases of epilepsy. The motivation for such a heavy and irreversible surgery lays in the interpretation of hard epileptic crises as an electro-magnetic storm spreading through all the brain from isolated sources of unrest (Purves et al., 2002). Removing the corpus callosum "*helps to confine the epileptic seizure to one side and tends to preserve consciousness during an attack* […] *and enables the patient to take precautionary or control measures at onset of a seizure*" (Sperry et al., 1969: 274).

Behavioural issues of commissurotomy have been extensively studied in medicine as well as in neuroscience since the re-introduction of the commissurotomy. The most surprising result of this surgery is the apparent lack of change with respect to daily life (Sperry et al., 1969: 275). One year is in general required for recovery from the surgery. Within two years, patients are able to return to school or go to work to such an extent that a conventional medical examination cannot reveal anything extraordinary in the behaviour of these patients (Kolb and Whishaw, 2003: 433). One needs specific cognitive tests to detect differences between commissurotomized patients and healthy subjects from the behavioural point of view and to show that in the split-brain, each hemisphere processes separately one half of the information processed by the brain as a whole, leading to the conclusion that "*in the split-brain, each hemisphere can be shown to have its own sensations, percepts, thoughts, and memories that are not accessible to the other hemisphere*" (Kolb and Whishaw, 2003: 433).

Standard tests allowing to discriminate between split-brain and normal patients aim to determine whether sensory information presented to only one hemisphere is at disposal of both or only of one hemisphere for motor action. For instance, a very simple test consists in asking a blindfolded individual to touch an object with the left hand and to find a similar object in a hidden collection with the right hand (Kolb and Whishaw, 2003:437). As the left hand is under the somesthesic and motor control of the lone right hemisphere and vice versa for the right hand, in the absence of inter-hemispheric communication, split-brain patients are unable to match the objects correctly. Although we are not interested in developing all the tests designed to capture behavioural specificities of split brain patients, let us consider in more detail the case of visual perception and recognition of objects.

The connectivity of the human visual system is crossed in a very specific manner. Roughly speaking, visual information of the left visual hemifield is computed by the right hemisphere and vice versa. Of course, since this crossing concerns the visual hemifield and not the eyes themselves, the structure of the central visual pathway is quite sophisticated. Optical signals coming from the left visual field are received by the right half part of the retina of each eye. Both of them are connected to the right lateral geniculate nucleus through the optic nerves and the optic chiasm. From there, the information is transmitted by the optic radiations to the layer IV of the right visual cortex, namely the right striate cortex.

However, in the case of normal agents, information stemming from a specific half visual field is available to both hemispheres by the way of two mechanisms. First of all, the corpus callosum connects both hemispheres together, tying up homologous areas of the left and the right hemispheres. Secondly, eyes keep scanning micro-movements permanently. These unconscious micro-movements have the effect of distributing a large part of the information stemming from the centre of the visual field to both sides of the visual system, which would otherwise be confined to a specific hemisphere. The existence of these scanning micro-movements explains why split-brain patients are able to live quite normally as regards their visual abilities in spite of the lack of communication between both hemispheres. Since information stemming from the central part of the visual field is sent to each side of the visual cortex, left sided visual information can for instance be employed to form a verbal report, although the main speech area is principally left-sided. Scanning micro-movements provide therefore a natural and very powerful mechanism of compensation that enables split-brain patients to partially overcome their deficit of communication between hemispheres in the domain of visual perception.

In the case of visual perception, classical tests run as follows (Sperry et al., 1969: 275-277). The subject is seated at a table and faces an adjustable screen that prevents her from seeing her hands. Images are flashed on one or the other half of the screen whilst the subject is asked to fix a red dot in the central part of the screen. Images are flashed during circa 1/10 seconds to prevent the scanning micro-movements of the eyes from transmitting information to both hemispheres. Visual material can in that manner be presented selectively to one specific or both hemispheres. Objects are placed on the table in the back of the screen in such a way that the patient can be asked to identify them just by touching them. The experimenter tries then to determine which visual information is at disposal of which hemisphere for motor action.

In the execution of such tasks, commissurotomized patients appear to be essentially normal with respect to the right half visual field under these testing conditions. They can describe, read and use visual material in the same manner as before the surgery. However, if an image is flashed on the left side of the screen in a randomized sequence of stimulations, verbal reports of patients tend to show that they do not perceive the stimulus. If the time duration of the presentation of the stimulus is increased, patients are able to report what is presented in the left visual hemifield, relying on the mentioned compensation mechanism. It should be emphasized that commissurotomized patients do not recover their pre-surgery abilities in carrying out such tasks. Such behavioural consequences of commissurotomy seem to be definitive.

Looking at the permanent behavioural effects of the commissurotomy, these results could at first glance also be taken to indicate a defect in the right visual system of the patients. But further tests exclude this possibility: when non verbal responses are employed to evaluate patients' perceptions, we can be sure that they understand the task and that they perceive and identify the stimulus. For example, if an image of an object is flashed on the left visual hemi-field, in spite of the pathological incapacity of verbally reporting the stimulus, patients are nonetheless able to pick out by touch with the left hand a corresponding object in a collection of other items. Further tests demonstrate that right sided visual abilities of these patients are clearly unimpaired by the commissurotomy. A defect in verbal abilities is as well excluded since the subjects can produce complex verbal reports about right-sided stimulations. The cause of the disconnection syndrome appears therefore not to be a defect in primary sensory areas and the motor cortical area, but a defect within the information processing between both of them and more specifically to be the consequence of a lack of communication between both hemispheres.

Let us sum this section up by emphasizing that a large variety of such visual tests confirm the conclusion that in the case of commissurotomized patients, visual information stemming exclusively from the left or the right half of the visual field is processed separately in each hemisphere. Normal interaction between elements within each half-field is preserved, but split brain patients cannot integrate information of both halves of the visual field. However, basic mechanisms allow these patients to lead a normal life, although their behaviour differs notably in many aspects once specific cognitive tests are employed. Using such cases as example, the next section of the paper is intended to make clear how the sub-type strategy of reduction can be applied to the relation between psychology and neuropsychology.

3. *Application of the sub-type strategy to the split-brain example*

Let us introduce the following example to get the application of the strategy of reduction by means of functionally defined sub-types started. Take the case of two pairs of individuals watching TV and discussing the content of a very attractive film. Each individual can be described in psychological terms as *perceiving* the TV screen and, from that point of view, each of them instantiates the same perceptual mental property. The first step of functional reduction consists in the functionalization of that mental property. The considered property can be functionalized as follows:

(A) A subject S visually perceives an object x if, when she looks conspicuously at x and when she desires to give a verbal description of x, S verbally describes x in a successful manner.

There are of course many other cases of perception, as common sense ascribes this type of mental state in a lot of other scenarios. For example, verbal report is not a necessary condition for perception. But our purpose here is not to provide an extensive functional definition of perception. A generic formulation of such functional descriptions can be given in the classical Ramsey-Lewis shape of theoretical concepts:

(F) $(\exists x)(\exists y)(\exists z)((x$ is caused by environmental conditions $c_1 \vee c_2 \vee \ldots \vee c_n) \;\&\; (x$ tends to cause mental states $y \vee z) \;\&\; (x$ tends to cause behaviour $b_1 \vee b_2 \vee \ldots \vee b_n))$

The symbol "∃" designates the existential quantification over a domain of individuals. Proposition (F) can be read in English as "There is a property token x, and there is a property token y, and there is a property token z, such that …".

The second step of functional reduction consists in the identification of physical truthmakers of (A). Let us now assume that one of the individuals in each pair is a split-brain patient. Nonetheless, (A) applies to both individuals in each pair. This is a clear case of multiple realization since one and the same functional description applies to individuals that come under two distinct neuropsychological descriptions. Let us also stipulate that the first pair of individuals is watching a perfectly normal movie and that the second pair is watching exactly the same movie, except that their version of the movie contains additional very short left-sided apparitions of a given item.

Since split-brain patients and normal individuals are neurophysiologically different, visual information is not processed by their respective brains in the same way. As mentioned in the preceding section, these differences can be detected behaviourally only in specific circumstances. Accordingly, the following situation is perfectly possible. Within the first pair, both the normal and the split-brain person behave in exactly the same way. Both of them perceive and comment the movie in exactly the same manner. However, within the second pair, the behavioural capacities of both individuals differ clearly. The normal subject perceives and comments the movie *being able to describe left-sided and short apparitions* that characterize the version of the film they are watching. The split-brain patient of that pair, however, perceives and comments the movie *without being able to describe left-sided and short apparitions*. Such behavioural differences are nothing but the side-effects we mentioned in section 1. They enable us to construct functional sub-types of the considered functional type by including behavioural differences in order to make our psychological descriptions more fine-grained. We thus get to the formulation of two new functional descriptions:

> (A') A subject *S* visually perceives a movie if, when she looks conspicuously at the movie and when she desires to give a verbal description of it, *S* verbally describes the movie in a successful manner *including possibly the description of left-sided apparitions of less than 1/10 second*.

> (A'') A subject *S* visually perceives a movie if, when she looks conspicuously at the movie and when she desires to give a verbal description of it, *S* verbally describes the movie in a successful manner *excluding systematically the description of left-sided apparitions of less than 1/10 second*.

The starting point was that the coarse-grained functional description (A) applies to all the individuals considered in our example. However, the more fine-grained description (A') applies in the second pair only to the normal individual, since the split-brain patient is unable to verbally report the short and left-sided apparitions. Consequently, the description (A'') by contrast to (A') applies to her.

Let us emphasize that this result is precisely what is required for applying the strategy of reduction by means of functional sub-types. Adding functional specifications to the general description at the outset has the consequence of reducing the extension of that description. Of course, description (A'') does not apply only to split-brain patients watching TV. There could be several other cases in which an individual is unable to notice the brief and left-sided apparitions. Nonetheless, we would be able to introduce some further specifications of the functional description (A'') in order to narrow down its extension in an even more fine-grained manner. For example, having distinct memories and motor abilities (Kolb and Whishaw, 2003: 433), the split-brain patient of the second pair is able to draw the object of the mentioned apparitions or to pick out a similar object in a hidden collection with the left hand, because this hand is under control of the right hemisphere, which is in possession of visual information relative to the apparitions. Sharpening in that manner the grain of our functional descriptions, it will be possible to capture precisely the behavioural particularities of a split-brain patient, namely e.g. the inability to successfully match sensory information originating solely from one side of the body with similar information originating from the other side. In other words, introducing functionally defined sub-types for both normal and split-brain patients enables us to reach the required condition for intertheoretic reduction, namely co-extensionality and thus bi-conditional bridging principles. Furthermore, neuroscience explains why the truthmaker of each functional sub-description produces a specific behaviour. We are thus able to meet the requirement of the third step of functional reduction.

Reductionism is often understood as implying the replacement or elimination of higher level generalizations by more precise and reliable generalizations at a lower level of description because through reduction the pattern of similarities brought out by the higher level generalizations is lost. However, the strategy of reduction by means of functionally defined sub-types has the resources to avoid such a loss. As the discussed example shows, definition (A) is preserved as a perfectly valid and well grounded high level (abstract) generalization by retaining functional specifications that are common to both definitions (A') and (A''). In other words, definition (A) stands in a relation of abstraction with respect to definitions (A') and (A''). The core of this abstraction consists in the pruning of the functional specifications that distinguish between the many realizers of a given higher level property. Introducing functional sub-types makes therefore room for a reductionist but conservative position with respect to the higher level generalizations of psychology, since the commonalities shared by the different realizers of a psychological property are characterized in a functional manner.

4. *Critical conditions of manifestation and the quest for the localization of functions*

The approach introduced in the preceding section depends crucially on the possibility of finding adequate testing conditions that bring out the functional side-effects discussed in section 1 in order to be able to discriminate from a purely behavioural point of view between the physically different truthmakers of a given mental description. What is required is that for each realizer p_n falling under the functional type F, there is at least one set of environmental conditions conceivable in which the specific way that p_n employs to bring about the effect characterizing F can be grasped from the psychological point of view, that is, by looking just at the behaviour. The question remaining at this stage is whether it is possible to find such environmental conditions for any specificity that is relevant from the point of view of neuropsychology, that is, for any neuropsychological type of realizer of a given functional type F.

A rather trivial way to answer this question consists in relying on the definition of neuropsychology given by Kolb and Whishaw, namely the "*study of the relation between* *human brain function and behaviour"* whose *"central focus is the development of a science of human behaviour based on the function of the human brain"* (2003: 2). From that point of view, neuropsychology is based on the fundamental premise that a systematic relation between brain specificities and behavioural specificities exists, and neuropsychological types are constructed in order to group under common descriptions different patterns of behavioural specificities that share a particular cause at the neurological level. It is then obvious that for any neuropsychological type realizing a given mental property, there are behavioural differences in the execution of some cognitive tasks.

A more sophisticated answer consists in looking in detail at the daily scientific practice. For instance, neuropsychologists typically try to locate the execution of cognitive functions within the brain. Focusing on that goal, their experiments are designed to create a statistically significant contrast between control subjects and pathological or medicated patients precisely in order to show the causal impact of neural differences on the production of behaviour. In other words, they aim to create a contrast between the following two pairs of correlations: on the one hand a correlation between the level of performance presented by control subjects in the execution of a given cognitive task and their intact neurological structures; on the other hand a correlation between the production of behavioural particularities by tested patients in the execution of the very same task and their neurological particularities.

The localization of cognitive functions has been pursued for a long time by comparing healthy brains and behaviour with pathological cases through a post-mortem investigation of the brain structure. The history of neuroscience contains several examples of that method, such as the double dissociation between the lexical function of Wernicke's area and the syntactical function of Broca's area. Damages to the Wernicke area cause impairment in language understanding and in the production of meaningful sequences of words. Damages to the Broca area tend to cause impairments with respect to the syntax and the fluency of speech and difficulties in the understanding of grammatically complex sentences. Such dissociations remain the basis for the localization of cognitive functions, albeit the recent development of modern neuroimaging multiplies the possibilities of investigation. Neuroimaging enables scientists to observe directly the neural activity and, in that manner, to establish a correlation between a specific result in the execution of a behavioural task and a particular neural activation or lack of activation.

In order to establish such correlations, the key issue is to design an experimental set up that asks the subjects to execute a certain task that comes as close as possible to the cognitive function that the cerebral area in question is presumed to implement. The difficulty for neuroscientists consists therefore, firstly, in making good hypotheses about cognitive functions, that is, in developing a model of brain information computation using the right taxonomy of functions and mechanisms (Craver, 2007: 128), and, secondly, in finding an experimental protocol that requires of subjects precisely the execution of that very function. Neuroscientists then are able to correlate behavioural results with patterns of neural activities.

Take as example of a complex neuropsychological pathology the case of Schneiderian schizophrenia and the recent developments in the identification of the neural mechanisms of recognizing self-generated actions. From the cognitive point of view, the central monitoring hypothesis constitutes the current general framework for understanding the mechanism of recognition of self-generated actions. Roughly, the idea is that "*each time the motor centres generate an outflow signal for producing a movement, a copy of this command ("the* *efference copy") is retained. The reafferent inflow signals are compared with the copy. If a mismatch arises between the two types of signals, new commands are generated until the actual outcome of the movement corresponds to the desired movement*" (Jeannerod, 2003: 8). Self-generated actions recognition is based in that framework on the concordance between a desired action and its predicted sensory consequences.

Defectiveness of theses mechanisms leads to cognitive impairments characterizing Schneiderian schizophrenia, such as auditory hallucinations, thought insertion, feeling of other's influence on the patient's thoughts, actions and emotions (Farrer et al., 2004: 31). For example, auditory hallucination such as voices hearing are explained within the framework of the central monitoring hypothesis as a pathological incapacity of the patient to recognise inner speech as her own (Franck et al., 2001). The motor areas are engaged offline in order to generate an inner linguistic representation (the inner speech), but the auditory cortex interprets this activity as if it were produced by an external cause (Jeannerod, 2003: 14). Schizophrenia appears that way as a pathology affecting the communication and the comparison of information between motors and sensory area.

In order to investigate the localization of the mechanisms of self-generated action recognition, many successive experimental designs have been set up (for a review, see Jeannerod, 2003). Farrer et al. (2004) designed an experiment whose central idea was to develop a task requiring of the subject to mobilise nothing but the execution of the investigated cognitive function, namely the comparison between a motor command and its intended result. The experimental set up used was one in which both healthy subjects and pathological patients had to judge whether or not the images displayed on a screen were effects of their own manipulations of an out-of-vision joystick. The experimenter was able to vary the degree of discrepancy between patient's inputs and visual output on the screen. Without entering in more details with respect to this experiment let us highlight two points.

First of all, brain imagery study of healthy subjects enables neuroscientists to correlate a task execution with a delimited pattern of neural activity. In the case of the studies of Farrer, a correlation has been reported between the degree of control by the subjects of the perceived movement on the screen and the activation of the right angular gyrus (Farrer et al., 2003; Farrer et al., 2004: 37), particularly of Brodman area 39 and 40 on the right side (reported in Jeannerod, 2003: 9). Maximal activation occurred when the movements shown to the subject were unrelated to the subject's own movements.

However, this correlation does not establish beyond doubt that the considered area is causally responsible for the correct task execution. It's only once one can compare, and that's here our second point, the results obtained with healthy subjects with results of impaired patients that the relation between neural activity and behaviour takes the form of a counterfactual dependency that makes it possible to consider the localized neural activity as a genuine cause of the observable behavioural particularity, which is precisely the result obtained by Farrer's research group. Schizophrenic patients perform the experimental task very poorly, showing that they encounter difficulties in distinguishing self-generated movements from alien generated movements, while exhibiting an aberrant activity of the right angular gyrus in the very same experimental conditions (Farrer et al., 2004: 37, 42).

This result confirms the hypothesis that the right angular gyrus is the locus of the comparison between efferent copy and reafferent signal (Jeannerod, 2003: 9) and, by the way, they contribute to the current explanation of the schizophrenia disease. Hallucinations and delusions are understood in terms of abnormal interactions between different cortical areas. This dysfunctional integration is explained at the cognitive level, in the framework of the central monitoring theory, as a failure to integrate perception and action; at the physiological level, it is considered to be a defect in connectivity (Farrer et al., 2004: 41). This abnormal connectivity "*disrupts the modulation by frontal region of those more posterior brain areas involved in the processing of the sensory consequences of an action*" (Farrer et al., 2004: 41), making it difficult to identify the source of the perception as internal or external. The incriminate absence of modulation then causes the primary sensory area relative to the perceptive modality in which the hallucination occurs to process the sensory consequence of action as if it were the result of an external cause (Blakemore et al., 2000, Frith and Dolan, 1996).

Let's return to the functionally defined sub-types strategy. The discussed example of complex neuropsychological pathology enables us to make the following remarks. First of all, a research programme like the one of Farrer relies on the possibility of finding critical testing conditions that provide an empirical feedback with respect to the hypothesis about the cognitive sub-functions that together compute the studied cognitive ability, namely here the set of sub-functions taken for granted by the central monitoring hypothesis and the ability of recognising self-generated actions. As said above, this possibility is open *ipso facto* in cases of schizophrenia and other neuropsychopathologies, since all of these concern cases in which the impairment is observable from the behavioural point of view.

Second, the cognitive tests used by neuropsychologists capture critical conditions that enable us to establish a contrast at the physiological level that mirrors the contrasted results obtained at the behavioural level. These critical test conditions are precisely the salient conditions required in order to construct functionally defined sub-types. In the case of schizophrenia, healthy and pathological subjects are clearly distinguishable by means of their results in the test used in the experiments of Farrer's research group. Since the results of these tests are graspable from a behavioural point of view, they can be employed in order to make the mental functional definitions more precise, as explained in the case of split-brain patients. Of course, the detailed functional reports at issue in the case of such a complex pathology as schizophrenia will be much more complex. However, this is a practical in contrast to a principled problem.

The requirements of the reduction strategy by means of functionally defined sub-types are thus met. The possibility of constructing sub-types exists for complex neuropathology with the result that one can obtain fine-grained psychological descriptions that are coextensive with neurological descriptions. In fact, neuroscience makes considerable progress in view of producing reductive explanations of the neural causes of pathologies such as schizophrenia. The strategy of reduction by means of functional sub-types makes it possible to receive this progress in a conservative rather than an eliminativist manner: the classifications in mental terms get their scientific quality vindicated by being systematically linked with neuropsychological classifications via mental, functional sub-types.

At this point, it should be clear that there is no principled objection to applying the procedure to any pathology within neuropsychology, that is to any pathology whose effects are observable from the behavioural point of view. Hence, it is in principle possible to introduce a functionally defined sub-type of a mental type for each type of a neuropsychological realizer, that is for normal individuals in contrast to split-brain or schizophrenian patients, but also for neglect patients, blind-sighted patients, and so on. What is required is just to sharpen in the following way the functional reports, using the behavioural testing tools of neuropsychology:

> (F') $(\exists x)(\exists y)(\exists z)$(($x$ is caused by environmental conditions c_1 v c_2 v … v c_n) & (x tends to cause mental states y v z to occur) & (x tends to cause behaviour b_1 v b_2 v … v b_n)) + further descriptions of actual causes + further descriptions of mental states implicated + further descriptions of actual behavioural effects of x.

This sharpening is possible because all neurological pathologies induce specific behavioural effects that neuropsychology can employ in order to build a global theory of the production of behaviour by the brain. Nonetheless, we should be able to go further by applying our procedure to distinguish between non-pathological but distinct realizers as well. Since, as outlined in section 1, differences in composition imply causal differences, we should be able to find critical situations where differentiations in the structure of normal brains are linked with specific cognitive abilities that manifest themselves in certain behavioural effects under specific circumstances. These behavioural differences can then be employed to construct more finely grained sub-descriptions of our traditional functional descriptions. This procedure can in principle be reproduced down to the physical description of the brain. Of course, functional descriptions matching physical differences have to be extremely detailed, but, again, that's a practical problem, not a principled one speaking against reduction.

5. *Conclusion*

As argued in this paper, we can achieve a conservative reduction of psychology to neuropsychology by means of functional sub-types by respecting the following guidelines. First of all, ordinary psychology has to be arranged in accordance with the standard functionalist interpretation of the special sciences, using common-sense functionalism or scientific functionalism. Mental properties are thus identified through their characteristic effects. Second, neuropsychology explains how a specific brain structure produces certain behaviour and why neurological variations lead to variations in the way of producing the behaviour in question. Third, we can construct functionally defined sub-types of mental types using standard cognitive tests in order to achieve coextension of sub-types, whose description is couched in the vocabulary of common-sense functionalism, and the corresponding neuropsychological descriptions. Those cognitive tests are designed to grasp functional side-effects produced by specific neuropsychological realizations of mental types. Ultimately, since any compositional differences imply causal differences, it should be possible to replicate this procedure in order to reduce neuropsychological descriptions to some more fine-grained neuroscientific descriptions, and, in the last resort, to a physical description of the brain at the molecular level.

References

Bechtel, W. & Mundale, J. (1999), 'Multiple realizability revisited: linking cognitive and neural states', *Philosophy of Science*, **66**, pp. 175-207.

Bedeschi, M. F., Bonaglia, M. C., Grasso, R., Pellegri, A., Garghentino, R. R., Battaglia, M. A., Panarisi, A. M., DiRocco, M., Ballotin, U., Bresolin, N., Bassi, M. T. & Boratti, R. (2006), 'Agenesis of the corpus callosum: clinical and genetic study in 63 young patients', *Pediatric Neurology*, **34**, pp. 186-94.

Bickle, J. (1998), *Psychoneural reduction: the new wave* (Cambridge, Massachusetts: MIT Press).

Blakemore, S. J., Smith, J., Steel, R., Johnston, E. C. & Frith, C. D. (2000), 'The perception of self-produced sensory stimuli in patients with auditory hallucination and passivity experiences: evidence for a breakdown in self-monitoring', *Psychological Medicine*, **30**, pp. 1131-39.

Craver, C. F. (2007), *Explaining the brain: mechanisms and the mosaic unity of neuroscience* (Oxford: Oxford University Press).

Daprati, E., Sirigu, A., Pradat-Diehl, P., Franck, N. & Jeannerod, M. (2000), 'Recognition of self-produced movement in a case of severe neglect', *Neurocase*, **6**, pp. 183-201.

Dobyns, W. B. (1996), 'Absence makes the search grow longer', *American Journal of Human Genetic*, **58**, pp. 7-16.

Esfeld, M. & Sachse, C. (2007), 'Theory reduction by means of functional sub-types', *International Studies in the Philosophy of Science*, **21**, pp. 1-17.

Farrer, C., Franck, N., Frith, C. D., Decety, J., Georgieff, N., D'Amato, T. & Jeannerod, M. (2004), 'Neural correlates of action attribution in schizophrenia', *Psychiatric research: Neuroimaging*, **131**, pp. 31-44.

Farrer, C., Franck, N., Georgieff, N., Frith, C. D., Decety, J. & Jeannerod, M. (2003), 'Modulating the experience of agency', *Neuroimage*, **18**, pp. 324-33.

Franck, N., Farrer, C., Georgieff, N., Marie-Cardine, M., Daléry, M., D'Amato, T. & Jeannerod, M. (2001), 'Defective recognition of one's own actions in patients with schizophrenia', *American Journal of Psychiatry*, **158**, pp. 454-59.

Frith, C. D. & Dolan, R. (1996), 'The role of the prefrontal cortex in higher cognitive functions', *Brain Research: Cognitive Brain Research*, **7**, pp. 1505-12.

Jeannerod, M. (2003), 'The mechanism of self-recognition in humans', *Behavioral Research*, **142**, pp. 1-15.

Kim, J. (1998), *Mind in a physical world. An essay on the mind-body problem and mental causation* (Cambridge, Massachusetts: MIT Press).

Kim, J. (1999), 'Making sense of emergence', *Philosophical Studies*, **95**, pp. 3-36.

Kim, J. (2005), *Physicalism, or something near enough* (Princeton: Princeton University Press).

Kolb, B. & Whishaw, I. (2003), *Fundamentals of human neuropsychology* (New York: Worth).

Lewis, D. (1980), 'Mad pain and Martian pain', in: N. Block (ed.), *Readings in the philosophy of psychology. Volume 1* (London: Methuen), pp. 216-22. Reprinted in D. Lewis (1983), *Philosophical papers. Volume 1* (Oxford: Oxford University Press), pp. 122-30.

Lewis, D. (1994), 'Lewis, David: Reduction of mind', in: S. H. Guttenplan (ed.), *A companion to the philosophy of mind* (Oxford: Blackwell), pp. 412-31. Reprinted in D. Lewis (1999), *Papers in metaphysics and epistemology* (Cambridge: Cambridge University Press), pp. 291-324.

Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., McMantia, A. S. & McNamara, J. O. (2002), *Neurosciences* (Paris: Bibliothèque nationale).

Sachse, C. (2007), *Reductionism in the philosophy of science* (Frankfurt, Main: Ontos).

Sirigu, A., Daprati, E., Pradat-Diehl, P., Franck, N. & Jeannerod, M. (1999), 'Perception of self-generated movement following left parietal lesion', *Brain*, **122**, pp. 1867-74.

Sperry, R. W., Gazzaniga, M. S. & Bogen, J. E. (1969), 'Interhemispheric relationship: the neocortical commissures; syndrome of hemisphere disconnection', in: P. J. Vinken & G. W. Bruyn (eds.), *Handbook of clinical neurology* (Amsterdam: North Holland), pp. 273-90.