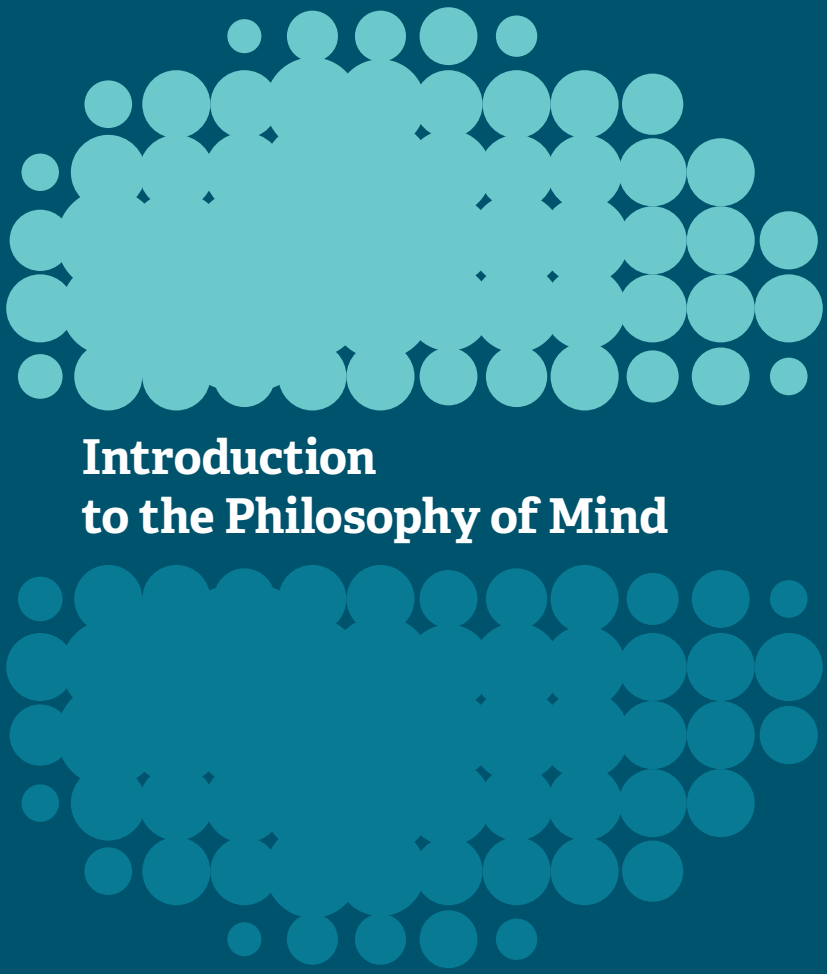

Introduction to the Philosophy of Mind

Silvia Gáliková

Edition Cognitive Studies

ffpu

Introduction to the Philosophy of Mind

Silvia Gáliková

Edition Cognitive Studies

fftu

The publication of this book is part of the project *Innovative Forms of Education in Transforming University Education* (code 26110230028) — preparation of a study program Cognitive Studies, which was supported by the European Union via its European Social Fund and by the Slovak Ministry of Education within the Operating Program Education. The text was prepared in the Centre of Cognitive Studies at the Department of Philosophy, Faculty of Philosophy in Trnava.

fftu

ISBN 978–83–7490–611–1

Contents

I. Philosophy and science of mind

Key words: *neodualism, subjective experience, phenomenal content, naturalism*

What is the real nature of mental states and processes? In what medium do they take place, and how are they related to the physical world? Will my consciousness survive the disintegration of my physical body? Or will it disappear forever as my brain ceases to function? Is it possible that a purely physical system such as a computer could be constructed so as to enjoy real conscious intelligence? Where do minds come from? What are they?

These are some of the questions we shall confront in the following text. Which answers we should give to them depends on which theory of mind proves to be the most reasonable theory on the evidence, to have the greatest explanatory power, predictive power, coherence, and simplicity. Let us examine the available theories, and the considerations that weigh for and against each.

The curiosity of Man, and the cunning of his Reason, have revealed much of what Nature held hidden. The structure of spacetime, the constitution of matter, the many forms of energy, the nature of life itself; all of these mysteries have become open books to us. To be sure, deep questions remain unanswered and revolutions await us still, but it is difficult to exaggerate the explosion in scientific understanding we humans have fashioned over the past 500 years. Despite this general advance, a central mystery remains largely a mystery: the nature of *conscious intelligence.* That

is what this book is about. If conscious intelligence were still totally mysterious, there would be no useful book for me to write. But encouraging progress has indeed been made. The phenomena to be penetrated are now the common focus of a variety of related fields. Philosophy has been joined by psychology, artificial intelligence, neuroscience, ethology, and evolutionary theory, to name the principals. All of these sciences have made contributions to what used to be a purely philosophical debate, and all of them promise much more to come. This aims at introducing some of the main elements of the current philosophical/scientific debate on the major issues, the competing theories and the most important arguments and evidence. In the past decades, philosophy has made significant progress on the nature of mind: mainly by unraveling the status of the mind's self– knowledge, but also by providing a clearer conception of the nature of the possible alternative theories of mind between which we must finally choose, and by clarifying what sorts of evidence are needed if we are to make a reasoned choice between them. More important still, the empirical sciences mentioned have provided a steady flow of evidence relevant to the making of such a rational choice. Psychology has taught us some surprising things about the penetration and reliability of our introspective knowledge. Cognitive psychology and artificial intelligence have produced provocative models of cognition, which, when 'brought to life' within a suitably programmed computer, mimic closely some of the complex activities of goal–driven intelligence. The neurosciences have begun to unravel the vast microsystem of interconnected brain cells that, in living creatures, appears to execute those activities. Ethology has given us new insights into the continuities, and discontinuities, relating human intelligence with the intelligence of other creatures. And evolutionary theory has revealed the long and intricate selective processes from which conscious intelligence has slowly emerged. The evidence is still ambiguous, however, and a choice from among the relevant theories has not yet been made, so the reader of this book will have the pleasure

and excitement of joining an intellectual adventure that is still very much in progress. What is the real nature of mental states and processes? In what medium do they take place, and how are they related to the physical world? With regard to the mind, these questions address what philosophers call the *ontological problem.* This problem is more widely known as the *mind–body problem,* and very probably you are already familiar with the most basic division in views here. On the one hand, there are *materialist* theories of mind, theories which claim that what we call mental states and processes are merely sophisticated states and processes of a complex physical system: the *brain.* On the other hand, there are *dualist* theories of mind, theories which claim that mental states and processes are not merely states and processes of a purely physical system, but constitute a distinct kind of phenomenon that is essentially nonphysical in nature. Many of us bring strong convictions to an issue such as this, and many will think that the choice between these alternatives is easy or obvious, but it is wise to keep an open mind here, whatever your convictions, at least until you have explored the lay of the land. There are at least five radically different versions of dualism, for example, and a comparable number of materialist theories, all very different from one another as well. There are not *two* theories here from which we must choose, but more like ten. It is therefore of high importance to lay them out and to try to evaluate the strengths and weaknesses of each. One important aspect of the mind–body problem is the *semantical problem.* Where do our ordinary common–sense terms for mental states get their *meaning*? What would count as an adequate definition or analysis of those special concepts that we apply to ourselves and to other creatures with conscious intelligence? One suggestion, perhaps the most plausible one, initially—is that one learns the meaning of a term like "pain" or "sensation of warmth" simply by attaching the relevant term to the relevant kind of mental state, as it is experienced in one's own case. But this view leads

to a number of problems, one of which may already have occurred to you at some point or other.

How can you be sure that the inner sensation, to which your friend has attached the term "pain", is qualitatively the same as the inner sensation to which *you* have attached that term? Perhaps your friend's inner state is radically different from yours, despite its being hooked up to behavior, speech, and causal circumstances in the very same ways it is hooked up in you. Your friend would thus behave in all respect as you do, despite the hidden internal difference. The problem is that this skeptical worry, once raised, seems impossible to settle, because it appears entirely impossible that anyone should ever have *direct* experience of someone *else's* mental states, and nothing less than such experience would settle the issue.

If this is so, then it appears that none of us knows, or can know, what meaning the many terms for mental states have for other people, if indeed they have any meaning at all. One can know only what meaning they have for oneself. This is a very odd conclusion to reach about a major segment of our language. The purpose of language, after all, is public communication within a shared network of understanding. A competing theory of meaning suggests a different source for the meaning of our ordinary psychological vocabulary. To learn the meaning of the term "pain", it is claimed, is to learn that pain is a state that is often caused by bodily damage, a state that in turn causes other inner states such as mild unhappiness or outright panic, a state that causes characteristic sorts of behavior such as wincing, nursing, and moaning. In short, the essential feature of pain is said to be a *network of causal relations* that connects any pain to a variety of other things, especially to publicly observable things.

Materialists of all persuasions tend to prefer this latter approach to meaning, partly because it leaves wide open the possibility that mental states are really physical states. There is no problem in supposing a purely physical state to have the appropriate kinds

of *causal* connections essential to being a pain. And this approach does not land us swiftly in skepticism. On the other hand, it does seem to give short shrift to the inner, introspectible aspect of our mental states, the aspect on which the first approach to meaning was centered. Dualists have tended to prefer that first approach to meaning, despite its apparently skeptical consequences. The introspectible or 'subjectively evident' qualities of our mental states represent for them some of the very essence of mentality, an essence that is beyond merely physical explanation.

These issues lead naturally enough to the *epistemological problem.* This problem has two parts to it, both very perplexing. The first arises swiftly from a worry already discussed: On what grounds has one the right to assume that other humans, for example, enjoy any mental states *at all*? Granted, the assumption that they do is one of the deepest assumptions one makes. But what exactly are the rational grounds for that assumption? To justify that assumption, what one needs to know is that the behavior of others is causally connected, in the same ways, to inner states of the same kind as those to which one's own behavior is connected. One needs to know, for example, that what is caused by a hammer blow and what causes in turn a loud "ouch!" is the *same* in others as in oneself. But that would seem again to require the impossible: the direct subjective experience of someone else's mental states. This is called the *problem of other minds,* and it is not merely a skeptical conundrum about our fellow humans. The problem begins to look less frivolous or academic when one starts to ask seriously after the mental life of animals like the great apes, or domestic dogs, or dolphins. Do they have genuine consciousness? And the current explosion in computer technology promises a new location for the problem. How can we distinguish a truly conscious intelligence from a complex physical system built to resemble a thinking being in all of its behavior, verbal and emotional behavior included? Would there *be* a difference? How could we tell? In sharp contrast to the opacity of the mental life of people other than oneself is

the transparency of one's own mental life. Each of us is self–conscious. What is the nature of that curious access you have to the contents of your own mind, but to no other? How is it you are able to tell, without looking at your behavior, what you feel, think, and desire? We take it for granted, this capacity for *introspection,* but it is a most extraordinary and enigmatic talent to have. A great deal has been claimed for it by various thinkers: infallibility, by some; the feature that distinguishes mind from matter, by others. And it does present a daunting challenge to any materialist who aspires to explain it. It is evident, that the nature of mind is not a purely philosophical question, but a deeply scientific question as well. To say this is not to beg any questions as to which of the alternative theories will be vindicated. But I do mean to assert that empirical research will weigh heavily, or even decisively, in determining the outcome. Which raises the question: What is the proper approach or methodology to pursue in constructing a 'science of the mind'? Here again there are differences. Should a science of conscious intelligence actively seek continuity with the network of established natural sciences (physics, chemistry, biology, and so on)? Or should it claim a discontinuous autonomy on grounds of some unique feature? What sorts of data should it admit as legitimate? Introspection? Behavior? Neurophysiology? These issues make up the *methodological problem,* and they are pointed toward the future.

Intelligence appears likely to be a fairly widespread phenomenon in the universe, and all advanced instances of it will inevitably face the problem of constructing a useful conception of just what intelligence *is.* That process of self– discovery, to judge from our own case, need not be an easy one. Neither will it be completed in a short period, if indeed it can ever be truly *completed.* But progress is still possible, here, as elsewhere in the human endeavor; and we must be prepared to contemplate revolutions in our conception of what *we* are, just as we have successfully navigated repeated revolutions in our conception of the universe that embeds us.

In everyday experience we intuitively feel that our intentions and decisions play an important role in action and behaviour. We feel that we cause ourselves to behave. So, it seems that we consciously will our voluntary actions. Experimental research and clinical practise however, bring evidence of fundamental flaws in the way people perceive themselves from the „inside", from the first person. nature of conscious states is *mental* as opposite to physical. These claims are supported by strong evidence from experimental research and clinical practise. Understanding subjective experience in naturalistic terms reveals how mind and consciousness fit perfectly into the physical world and therefore really *matter.*

Suggested readings

Gáliková, S.: *Psyché.* Bratislava: Veda, 2007.
Gáliková, S., Gál, E.: *Antológia filozofie mysle.* Bratislava: Kalligram, 2003.
Churchland, P. S.: *Brain–Wise: Studies in Neurophilosophy*. Cambridge: MIT Press, 2002.

II. Mental content of representations

Key words: *propositional attitudes, mental content, causal role, teleology, funcional role*

In philosophy of mind human thoughts, decisions and beliefs are considered as *propositional attitudes* — by this term philosophers mean either mental states, language (sentences) about mental states or their outcome, for exmaple in the form of a mental act. State, which is the content of a particular mental states has been labelled — a *proposition*. According to B. Russell, a mental states with a content are *propositional attitudes,* mostly in the form of a sentence with a „that" clause following the verb. For example, Mary's belief that the carpet is two meters wide, is about the carpet, represents the carpet which is two meters wide. In connection with propositional attitudes, a great deal of attention has been paid to the issue of **wide** versus **narrow content.** The wide/narrow distinction was originally drawn by the contemporary American philosopher Hilary Putnam, although Putnam was primarily concerned with the meaning of linguistic items rather than mental content (Putnam, 1975). In what follows we will focus on the wide/narrow distinction as it applies to mental content. Here's an example much like that which Putnam used. Peter frequently has thoughts about water, and since water is H_20, when Peter has a water thought he's thinking about H_20. Now, in a distant part of the universe there is a planet that is almost exactly the same as Earth, and living on that planet is a guy who's almost exactly like Peter. Let's call that planet **Twin–Earth** and call the guy who's almost exactly like Peter 'Twin–Peter'. I said that Twin–Earth is almost exactly like Earth, and that Twin–Peter is almost exactly like Peter. In fact, with one small exception, Twin–Earth is an atom–for–atom duplicate of Earth, and Twin–Peter is an atom–for–atom duplicate of Peter. The exception is this: whereas here on Earth the clear liquid which fills lakes, comes out of taps, and is essential for life is H_20, on Twin–Earth the clear liquid which fills lakes, comes out of taps, and is essential for life is XYZ. (Or, as I will say henceforth, whilst the 'wet stuff' on Earth is H_20, the 'wet stuff' on Twin– Earth is XYZ.) Moreover, in all superficial respects H_20 is strikingly similar to XYZ—so much so that without a chemical analysis you can't tell them apart. Now let's consider Twin–Peter's thoughts about the wet stuff on Twin–Earth. Peter's thoughts about the wet stuff on Earth are about H_20; it's implausible, though, that Twin–Peter is thinking about H_20. After all, he's never seen, touched, or drunk any H_20, and he might not even know that it exists. Rather, when Twin–Peter thinks about the wet stuff on Twin–Earth he's thinking about XYZ. So the content of Peter's thoughts about the wet stuff is distinct from that of Twin–Peter's thoughts about the wet stuff. Peter's thought is about H_20; Twin–Peter's thought is about XYZ.What's striking about this case is that, even though Peter and Twin–Peter's have brains which are in all relevant respects identical, their thoughts about the wet stuff have different contents. It follows that the content of our thought is not entirely determined by our brain states: it's possible for two people to have identical brain states and yet have thoughts with different contents. To use an old slogan: "Meanings ain't in the head" (Putnam 1975). With this example belore us, we can appreciate the difference between wide and narrow content. Philospohers say that Peter's beliefs about wet stuff have different wide content to Twin Peter's beliefs about the wet stuff. The contents of their beliefs are wide in that they are **individuated** (or distinguished) by what's going on in the world external to the believers' heads. Peter and Twin–Peter's thoughts about the wet stuff are

distinct, not because their brains are in relevant respects distinct, but because their *environments* are distinct: one's in an H_20 — containing environment, the other an XYZ –containing environment. Whilst we can recognize a sense in which the content of Peter's belief differs from that of Twin–Peter's, there's also a sense in which the contents of their beliefs are the *same.* Say that Peter believes that he should drink eight glasses of water every day, and that he expresses his belief by saying, 'I should drink eight glasses of water every day'. Then Twin–Peter will have a belief which he expresses with the phrase, 'I should drink eight glasses of water every day'. Moreover, note that Peter's belief will cause him to behave in certain ways: other things being equal, he'll drink eight glasses ol the stuff he calls 'water' every day. Similarly, Twin–Peter's belief will cause him to act in certain ways: other things being equal, he'll drink eight glasses of the stuff he calls 'water' every day. Finally, Peter and Twin–Peter will both feel the same way about the stuff they call 'water'. If Peter has a phobia about washing in the stuff he calls 'water', then Twin–Peter will have a phobia about washing in the stuff *he* calls water.

So far we have recognized that in many ways Bloggs and Twin–Bloggs would seem to have identical beliefs about the wet stuff. But identity of belief implies identity of content. So if there's a sense in which Bloggs and Twin–Bloggs have the same beliefs about the wet stuff, then there's a sense in which their beliefs about the wet stuff have the same content. The expression *narrow content* is used to pick out the content which Bloggs and Twin–Bloggs share.The focus of our discussion so far has been Putnam's famous Twin–Earth example. The contemporary American philosopher Tyler Burge (1979) has provided another kind of example of wide content. Say that Peter wakes up one morning with a pain in his leg, halfway between his knee and his hip. 'Goodness,' thinks Peter, 'I've got arthritis in my thigh.' In fact, arthritis is, by definition, inflammation of a *joint,* so Peter can't have arthritis in his thigh. Consequently, Peter's belief that he has arthritis in his thigh is false. Now consider

a slightly different situation. Imagine that the medical profession uses 'arthritis' not for inflammation of the joints but for leg pain; that is, imagine that 'arthritis' means 'pain in the leg'. Everything else about the situation remains as before; in particular, Peter's brain states are exactly as they were before. But now when Peter wakes up in the morning and thinks, 'Goodness, I've got rthritis in my thigh', he has a *true* belief. He really does have arthritis since he has a pain in his thigh and 'arthritis' means 'pain in the leg'.It's clear that the content of Peter's beliefs about arthritis are determined by facts outside Peter's head; in particular, they're determined by facts about the way the word 'arthritis' is used in the broader community. So once again we have a case of wide content: meaning ain't in the head. In everyday life we typically pick out beliefs by their wide content; that is, we identify beliefs in terms of objects external to the believer. However, there's a strong case to be made for arguing that scientific psychology should distinguish beliefs in terms of their *narrow* content. Imagine that Peter is transported to Twin–Earth and interpret the content of his desire to drink eight glasses of water per day narrowly. Relying on the principle that, other things being equal, people act so as to satisfy their desires, we can predict that when he is on Twin–Earth Peter will drink eight glasses of the wet stuff per day; that is, he will drink eight glasses of XYZ. That sounds like the right prediction to make—after all, he won't be drinking eight glasses of H_20 per day because there's no H_20 for him to drink. Now imagine that Peter is transported to Twin–Earth and interpret the content of his desire to drink eight glasses of water per day widely. Understood widely, his desire to drink eight glasses of water per day is the desire to drink eight glasses of H_20 per day. If we now apply the principle that, other things being equal, people act so as to satisfy their desires, we end up predicting that on Twin–Earth Peter will drink eight glasses of H_20 per day. And that's got to be wrong since, as we have noted, there's no H_20 for him to drink. Since scientific psychology aims at predicting behavior, it seems that it should individuate beliefs by

their narrow content. Even if this seems to be a standard consclusion, not everyone is ready to abandon wide content for predictive purposes. The problematic situation addresses the following ideas and claims: 1. Some mental stales have content; that is, they are *about* things. 2. Theories of content attempt to explain how mental states get to be about things. 3. According to the resemblance theory of content, mental states are about what they resemble. This theory faces severe difficulties. For the causal theory of content, dog thoughts are about dogs because they are caused by, and only by, dogs. The most widely discussed difficulty for any causal theory is the *disjunction problem* which arises because, in cases of misidentification, dog thoughts are caused by non–dogs—for example, by sheep. In that case the causal theory is committed to the claim that dog thoughts are about dogs–or–sheep. In the next chapter we will see how the question of mental content relates to the problem of the nature of mind.

Suggested readings

Crane, T.: Concepts in Perception. In: *Analysis* 48 (June): 150–53, 1988.

Evans, G.: *The Varieties of Reference*. Oxford: Oxford University Press, 1982.

Putnam, H.: *Mind, Language and Reality. Philosophical Papers, vol. 2*. Cambridge: Cambridge University Press, 1975.

III. Mental versus physical causation

Key words: *anomalous monism, causal exclusion, overdetermination, category mistake*

On of the well known phrases in philosophy of mind is Jerry Fodor's suggestion, that if it isn't literally true that my wanting is causally responsible for my reaching, and my itching is causally responsible for my scratching, and my believing is causally responsible for my saying, then practically everything I believe about anything is false and it's the end of the world. In these words Fodor expressed the common view about the indubitable existence of mental causality. His ideas address a list of properties which any theory of mental states has to explain or explain away. Three items on that list were concerned with the causal relations in which some mental states are typically involved: a) they are caused by states of the world; b) they cause actions and c) they cause other mental states. The task of explaining how mental states can have these kinds of properties is often called the *problem of mental causation*. Other things being equal, a theory of mental states which makes sense of mental causation is preferable to one which doesn't. For example, Descartes had difficulty explaining how the nonphysical mind he postulated interacts with the physical body. Descartes made two central claims: (i) that mind and body are radically different kinds of substances and (ii) that mind and body causally interact. As Princess Elizabeth of Bohemia pointed out, these two claims are in tension: how can two radically distinct kinds of substance

causally interact? When it comes to mental causation it might seem that physicalist theories have the upper hand: if mental states are physical states, then surely there can be no problem accounting for mental causation. If only it were that simple. In the 1980s philosophers began to realize that physicalism has its own problem of mental causation. Indeed, the contemporary American philosopher Jaegwon Kim has pointed out that physicalism laces *several* problems of mental causation (Kim, 1998). In the following text we will consider two problems: a) what Kim has called the *problem of causal exclusion* and b) a problem about the causal efficacy of content. In what follows I will talk about both mental states and brain states. When philosophers talk about states they are usually talking about *types*. For example, when they say, 'Other things being equal, fear causes screaming', they are making a claim about the type fear; they are only indirectly making a claim about particular tokens of fear. Kim uses expressions like "fear" and "brain state B" in the sense of types, and I will use expressions like "Peter's token of fear" when talking about tokens. **The problem of causal exclusion Peter** goes to see a horror movie, gets really frightened, and screams. Now, according to physicalism, mental states supervene on physical states; in particular, they supervene on brain stales. Let's say that, in Peter's case, his fear is realized by a token of brain state B. It is very likely that it is his being in brain state B that causes Peter to scream; that is, if we trace the causal pathway backwards from Peter's screaming we will, in due course, arrive at brain state B. But now we have a serious problem. What makes Peter scream is his being in state B; Peter's fear does nothing. But that's deeply counterintuitive: surely Peter screamed *because* he was afraid. Moreover, we posit mental states like fear in order to explain behavior. We attribute fear to Peter because, under the circumstances, attributing fear to him provides a good explanation of why he screamed. But if Peter's fear did not make him scream, then we no longer seem to have much reason for saying that he is afraid. If mental states don't do any causal work, there's little point in

positing them. This is the problem of causal exclusion. Once we allow that mental states are realized by physical states, there no longer seems to be anything for the mental states to do: they are 'excluded' from the causal story. One way to respond to the problem of causal exclusion is by allowing that Peter's fear *and* his being in state B caused him to scream. — an example of what philosophers call *overdetermination.* Imagine that Sally tells Petera joke and, simultaneously, Peter sees Neil slip on a banana skin. Both of these events cause Peter to laugh. Moreover, the events are independent of each other in this sense: if Peter had only heard the joke he would have laughed; and if he had only seen Neil slip he would have laughed. In that case Peter's laughing is overdetermined: it is independently caused by both the joke and the slipping. Whilst there are no doubt real cases of overdetermination, it's very hard to believe that every human action is overdetermined. It's simply incredible that everything I do is caused twice over: once by my mental state tokens and once by my brain state tokens. An alternative response would be to insist on the causal efficacy of Peter's fear whilst denying the causal efficacy of state B. But this, too, is unattractive. For whilst it preserves the intuition that Peter screamed because he was afraid, it denies the overwhelmingly plausible claim that a state of Peter's brain is causally responsible for his screaming. Notice that, if the identity theory is true, the problem of causal exclusion does not arise. For say that fear is type identical to brain state B. As it is one and the same state, there is no longer a threat of overdetermination. However, once we allow that fear is multiply realized, this solution is not available to us: it simply won't be the case that fear is type identical to a single physical type. Consequently, since most contemporary physicalists endorse the claim that mental states are multiply realizable, most contemporary physicalists are faced with the problem of causal exclusion. An ideal solution to the problem of causal exclusion would somehow maintain the causal significance of both mental states and the physical states which realize them, without falling prey to

overdetermination. A number of different responses to the problem of causal exclusion have been made. Broadly speaking, they fall into two categories. Those in the first category stress that analogous problems arise in cases in which there can be no serious doubt about the causal efficacy of the states involved. Those in the second category pay careful attention to the identification of tokens of mental states with tokens of brain states. Whatever you think about the claim that mental states are multiply realized, it's apparent that a great many properties which we encounter in the actual world are multiply realized. The property of being a stove, for example, is realized by a variety of physical states, as is the property of being a thermostat. The prevalence of properties which are multiply realized means that the problem of mental causation threatens to generalize. For example, the term "analgesic" is used to refer to those drugs which reduce pain. There are many different kinds of analgesic, including aspirin, paracetamol, and morphine. These substances differ in their chemical properties, and in the way they affect the nervous system. In other words, the property of being an analgesic is multiply realized. Now imagine that Bloggs has a headache and decides to take an analgesic. Twenty minutes later his headache is gone. As a matter of fact he took two aspirins, and aspirin causes certain changes in the brain which reduce pain. It seems that all the causal work is being done by the aspirin; the property of being an analgesic is causally inefficacious. It is therefore not only mental causation which is in trouble; by parity of reasoning, *any* property which is multiply realized is causally inefficacious. But this is absurd. It is a deep feature of our explanatory practices that we attribute causal powers to states and properties which are multiply realized. Rather than abandon those practices, we should stop worrying about the metaphysical problems to which they give rise. There must be *something* wrong with the arguments which generated the problems. This is the "Me, worry?" response to the causal exclusion problem. It's very implausible that analgesics have nothing to do with pain relief, that stoves

have nothing to do with saucepans boiling over, and that thermostats have nothing to do with shutting down furnaces. And it's also very implausible that my wanting a coffee has nothing to do with my going to the kitchen. It would be ridiculous to abandon causal explanations couched in terms of multiply realized properties. J. Kim (1998: 61–2) rejects the "Me, worry?" response. Philosophers want to know how the world "hangs together" in the most general sense; they want to know how the different parts or aspects ol the world relate to one another. The problem of causal exclusion is a classic philosophical problem because it suggests that our understanding of the world is inadequate. We all accept that my desire for a coffee caused me to go into the kitchen; and we all have good reason to accept the account of "kitchen–directed behavior" offered by the physiologist. But these two claims are in tension, and as philosophers we want to know why. The token identity theorist accepts that Peter's fear token is identical to a token of property B in Peter's brain. Now physiology has revealed that, other things being equal, tokens of property B cause tokens of screaming. Since the token of fear is identical to—*is one and the same thing as*—the token of B, neither the issue of overdetermination nor the issue of causal exclusion arises. So far we have only resolved the issue of the causal efficacy of *tokens* of mental states; we've said nothing about the issue of mental states understood to be *types.* We saw that if mental states are type identical to brain states, then the problem of the causal exclusion of mental states doesn't arise. However, according to the token identity theory, fear is not type identical to brain state B; consequently, the problem of causal exclusion, considered as a problem about types, remains. And the same applies to the old mind/body problem. Recently, there is a growing number of novel approaches and theories of the nature of mental content. In spite of prevailing methodological and theoretical problems philosophers and scientists aim together towards the explanation and understanding of our minds.

Suggested readings

Davidson, D.: Mentálne udalosti. In: Čin, myseľ, jazyk. Bratislava: Archa, 1997.
Kim, J.: „What is ‚Naturalized Epistemology'?", *Philosophical Perspectives*, Vol. 2, 1988, 381–405.
Ryle, G.: *The Concept of Mind.* London: Hutchinson, 1949.
Searle, J.: *The Rediscovery of Mind.* Cambridge: MIT Press, 1992.

IV. Minds in trouble

Key words: *seizure, episodes, consciousness, epilepsy, impairment*

Consciousness is an essential core feature of human life, so it is not surprising that disorders of consciousness have a major impact. Recent work has demonstrated that epilepsy shares many features with other disorders of consciousness. When patients lose consciousness during epileptic seizures they exhibit no meaningful responses to external stimuli; however, the eyes are usually open. In addition, although there is insufficient time to determine whether sleep–wake cycles are present, patients may exhibit orienting responses or other simple behaviors. Therefore, impaired consciousness during seizures resembles other disorders of consciousness such as vegetative state or minimally conscious state and, less so, coma. The major difference from these other disorders of consciousness is that seizures typically last for minutes rather than days, months, or years. The transient nature of epileptic seizures provides a unique opportunity for determining the anatomic and physiologic basis of impaired consciousness and its recovery. Seizures and other disorders of consciousness converge on a common set of cortical and subcortical structures. These structures constitute the "consciousness system," defined here as the bilateral medial and lateral fronto–parietal association cortex and subcortical arousal systems. Recent neuroimaging, intracranial EEG, and animal models demonstrate that the consciousness system forms a common anatomical substrate for all seizure types causing

impaired consciousness. The main types of seizures causing transiently impaired consciousness include absence seizures, generalized tonic–clonic seizures, and temporal lobe complex partial seizures. I will introduce the discussion on clinical and behavioral features, as well as recent neuroimaging and electrophysiology studies which have begun to shed light on the pathophysiology of impaired consciousness in these seizure types. It has been useful to separate consciousness into systems that are important for the *content* of consciousness, and those that control the *level* of consciousness. The content of consciousness can be thought of as the subject matter or substrate on which systems controlling the level of consciousness act. Thus, the content of consciousness includes all the information encoded in our hierarchically organized sensory and motor systems, as well as in the systems dedicated to memory and emotions. The level of consciousness is controlled by a specialized system of cortical and subcortical structures, which regulate alertness, attention, and awareness. Behavioral changes during absence seizures consist of arrest of ongoing movements, and lack of response to questions and commands. Absence seizures appear as if someone has "pushed the pause button" on the patient's behavior and responsiveness. Episodes are commonly accompanied by minor eyelid, mouth, or finger movements, but more significant motor activity is not part of typical absence seizures. The eyelids may droop but remain open; in fact if the eyes are closed before onset they tend to open during seizures. A few aspects of the EEG signal have been investigated in relation to behavioral impairment during absence seizures. First, as already noted, some studies report that longer EEG duration is related to more severe behavioral impairment in absence seizures. Neuroimaging of absence seizures has been greatly facilitated in the past 10 years by use of functional magnetic resonance imaging (fMRI), which has better spatial and temporal resolution than methods used in earlier studies. FMRI studies have revealed a complex sequence of fMRI changes in absence seizures including the main structures of

the consciousness system. The thalamus shows mainly fMRI increases, the medial frontoparietal and lateral parietal cortex mainly decreases, and lateral frontal cortex biphasic changes. Overall, the behavioral impairment and anatomic regions involved in absence seizures are similar to other more chronic disorders of consciousness. The scalp EEG during generalized tonic–clonic seizures shows high–frequency polyspike activity during the tonic phase, which gives way to rhythmic polyspike and slow–wave activity in the clonic phase. Postictally, while patients usually lie flaccid and unresponsive, the EEG shows generalized suppression, consisting of relatively low–amplitude EEG activity. Of interest, intracranial EEG recordings have shown that generalized tonic–clonic seizures do not involve the whole brain, and that some regions can be relatively spared. Neuroimaging of generalized tonic–clonic seizures cannot readily be done with fMRI because convulsions require close clinical attention and induce significant movement artifacts. Instead, insights have been gained from ictal single–photon emission computed tomography (SPECT) as well as positron emission tomography (PET). PET cerebral blood flow imaging also requires imaging during the convulsion. With SPECT, on the other hand, injection of radiopharmaceutical during the seizure is taken up within 20 to 30 seconds by the brain, providing a map of relative blood flow at the time of the injection. Therefore, imaging can be done later when the patient is clinically stable and no longer moving. Ictal SPECT is analyzed by comparison with baseline interictal SPECT in the same patients. Electroconvulsive therapy–induced seizures have the advantage of controlled timing and relatively consistent seizure onset. Seizures are induced by placement of electrodes in fixed locations either in the bilateral frontotemporal, bilateral frontal, or right unilateral regions. In summary, generalized tonic–clonic seizures usually cause complete unresponsiveness but the eyes are open, making the behavior resemble a transient vegetative state. Anatomic involvement of the consciousness system includes abnormal increased activity in the upper

brainstem and diencephalon, decreases in the medial frontal and cingulate cortex, and increases in the lateral frontal and mediolateral parietal association cortex. Postictal depressed cortical function may have a functional relationship with increased activity in the cerebellum. Further investigations are needed to better understand the mechanisms of selective network involvement in generalized tonic–clonic seizures, and the cortical–subcortical interactions governing postictal suppression of physiology and behavior. Impaired consciousness is classically seen in disorders that involve bilateral cortical–subcortical networks. The fundamental changes occurring in neocortical, subcortical, and limbic networks have recently been investigated in rodent models of hippocampal seizures. As in human temporal lobe epilepsy, rats with spontaneous limbic seizures following pilocarpine–induced status epilepticus as well as acute seizures induced by hippocampal stimulation exhibit fast activity in the hippocampus but slow 1– to 3–Hz activity in the frontoparietal cortex, associated with behavioral arrest. Impaired consciousness has a major negative impact on quality of life in patients with epilepsy. In large patient series the major factors that determined impaired quality of life in epilepsy were frequency and severity of seizures. Finally, it will be important to study behavioral interventions that may increase awareness by patients and families of impaired consciousness during seizures. This approach could help in the development of practical strategies for improving quality of life. In addition to impaired consciousness during seizures, patients with epilepsy commonly are unaware of the fact that they have had seizures and tend to underreport them. Recent human neuroimaging studies, intracranial EEG analysis, and animal model investigations have greatly increased our understanding of the fundamental mechanisms of impaired consciousness in epilepsy. The 3 seizure types causing impaired consciousness, namely absence, generalized tonic–clonic, and complex partial seizures, all converge on a final common set of anatomic structures we refer to as the consciousness system, consisting of medial and

lateral frontoparietal association cortex and subcortical activating networks. These same anatomic structures in the consciousness system are also affected in other states of impaired consciousness, including sleep, anesthesia, coma, vegetative state, and minimally conscious state. In behavioral terms absence or complex partial seizures often resemble a transient vegetative state, in which patients exhibit no meaningful behavioral responses, yet have open eyes and maintain some rudimentary postural tone and orienting responses. Other absence or complex partial seizures more closely resemble a transient minimally conscious state, because patients may show automatisms or variable simple responses yet do not demonstrate consistent functional interactive communication or object use. Improved treatments are needed to prevent impaired consciousness in epilepsy, particularly for patients in whom seizures cannot be stopped. Advances in understanding the fundamental mechanisms of impaired consciousness in epilepsy will be crucial for developing novel treatments targeting this major source of patient disability. Further work in this field will, it is hoped, lead to medications, surgical procedures, and behavioral interventions to reduce impaired consciousness and greatly improve the quality of life of patients with epilepsy.

Suggested readings

Eysenk, M., Keane, M.: *Kognitivní psychologie*. Praha: Academia, 2008.

Gazanniga, M., Ivry, R., Mangun, G.: *Cognitive Neuroscience: The Biology of the Mind*. New York: W.W. Norton, 2002.

Sacks, O.: *Muž, který si pletl manželku s kloboukem*. Praha: Mladá fronta, 1993.

V. Illusions of conscious will

Keywords: *illusion, consciousness, mental causality, experience, free will, mind, brain, first–person, third–person*

When using the concept of consciousness in everday experience we refer to „awareness", „wakefulness", „thinking" or „self–consciousness". It seems natural to consider states of consciousness as causes of our actions. For centuries we have pictured our actions as based on rationality, conscious will and responsibility. Contemporary experimental research, novel models and theories of mind unfold the ideal nature of this picture. In the following text I intend to refer to the hybrid nature of the concept of consciousness and the consequences for the study of the problem of explaining the nature of conscious experience. I will be concerned mainly with the concept of *phenomenal consciousness* which stays very close to everyday usage and it provides a simple place of departure on which widely diverging theories can agree. I will introduce examples of asymetry between introspectively experienced conscious states and a theortically founded explanation of ongoing causal processess. Neurocognitive research supports the existence of disproportions in the apparent mental causal chain of causes and effects. As recent findings show, experienced priority of thoughts before the act, consistency of thoughts about the planned act features more magical than causal relations in explaining person's behaviour. As Wegner put it conscisely „the experience of will is like magic" (Wegner, 2002, 489). Study of normal and impaired consciousness indicates that the

experience of conscious volitional activity can occur: a) before the act, b) right after the act, c) during the act. I appeal to the experimental research of D. M. Wegner (2002) on mutual relations between thinking, experiencing, timing and feeling responsibility of performed action. Person may feel conscious will in action he has not anticipated (confabulation of intentions) or he does not feel responsibility for performed action („allien–hand syndrom"). Besides that, when actions are caused unconsciously people tend to explain their behaviour in terms of mysterious forces (automatisms) or just „make up" stories (cognitive dissonance). Experimental findings demonstrate also as fundamental the *timing* problem in the supposed mental causality chain (Gray). Several experiments bring evidence for the delay of conscious states in a variety of modalities (motor action, language, thought). Automatic neuronal activity (ongoing objective time) preceeds reports of subjects (subjective time) about the *experienced will* to perform a motor activity (Libet, Wegner). In the following text I argue towards a profound asymetry between experienced introspective access to one's „inner world" and a theoretically adequate explanation of the nature of experience itself. In theory, similarly as in everyday experience, the concept of conscioussness has been used in a variety of meanings. I refer to the concept of consciousness in the sense of *subjectively experienced* (inner) life. In spite of much controveries over the concept of consciousness philosophers seem to agree that experienced inner states are inseparable from the explanation of consciousness. However, we can define several concepts that refer to consciousness as the experienced subjective life, for example: a) phenomenal consciousness, subjective experience as such, necessarily involving „qualia" or the qualitative character of experience; b) states and contents of consciousness which refer to the background state that allows specific contents of experience to appear in our minds; c) structure of phenomenal consciousness with clearest experiences in the center of consciousness; d) reflective consciousness, where we formulate conscious thoughts about other experiences

and thoughts; e) self–awareness, through which we „communicate" experiences to ourselves and others; f) unconscious information which could become conscious; g) zombies, a metaphor for complex, intelligent behavioral systems that operate in total absence of consciousness (Revonsuo, 2010, 96). Heterogenity of „definitions" of consciousness makes it hard to identify usages of the concept which are less confused and useful for the scientific understanding of consciousness. Unsuprisingly, differences in definitions of consciousness lead to contrary perspectives in investigating the nature and function of conscious states. It can be said that the widespread ambiguity of usage of consciouseness is also due to the complexity of the phenomenon itself. During past decades scholars studying normal and impaired states of consciousness have realized difficulty in grasphing variety of types of conscious states (visual, auditory, tactile, self–consciousness, etc.). Perhaps that is why philosophers and scientists recently prefer to consider consciousness as a continuum, with states of low attention on the one side (lethargy) and high attentative states (trances) on the other. Geunine and everlasting problem in consciousness studies is the problem of reconciliating first–person and third–person perspectives in explaning the nature of conscious states. Moreover, traditional dichotomies conscious–unconscious, inner–outer, subjective–objective appear in new forms and types, for example causal–phenomenal states of consciousness, access–phenomenal consciousness, state consciousness–creature consciousness etc. It seems that these distinctions serve as a so called „battle ground" in contemporary debates over the status of consciousness. Representatives of higher order thought theory of consciousness (HOT) „stand against" representatives of higher order perception theory of consciousness (HOP), followers of access consciousness (Block) „stand against" followers of phenomenal consciousness (Chalmers) etc. Eliminative materialists, behaviorists or functionalists, are automatically blaimed for rejecting the very existence of the phenomenon of consciousness. But i tis clear, that in reality, *nobody*

really wants to reject it! Since reductionists and anti–reductionists adopt fundamentally differing assumptions about the *ontology* of consciousness, they may intruded into how phenomenal consciousness has been defined. It is common for example for reductive physicalists and functionalists to take it for granted that an advanced form of brain science will ultimately demonstrate phenomenal consciousness to be nothing more than a state or function of the brain (Churchland, Dennett). If so, nothing would be lost by defining it in that way. However most theories of consciousness that resist a *reduction* of conscious phenomenology to brain states and/or functions fully accept that there is an *intimate relationship* between consciousness and brain (Searle). What is at stake is the *nature* of this intimate relationship. In his ulmost unrevised philosophical conception, J. Searle (1992) accepts that conscious states have special phenomenal properties, for example that they are intentional, subjective and private — all characteristics that traditionally distinguish the mental from the physical. However, as critics rightly point out, he then simply *declares* such facts about consciousness to be „objective *physical* facts" about the brain, thereby reducing the domain of the „mental" to a subclass of what is „physical" *by an act of redefinition* — leaving the problem of how objects such as brains could produce such intentional, subjective, private states untouched. Interpretations of many theoretical conceptions and approaches are made difficult by the fact that physicalist, functionalist, naturalistic dualist and modern dual–aspect theories agree that, in humans, every distinct conscious experience is likely to be accompanied by correlated activity in the brain (the neural correlates of consciousness). At the same time, naturalistic dualist, and dual–aspect theories resist the reduction of phenomenal consciousness to brain states (Chalmers). Dual–aspect theory suggests that conscious experiences and their correlated brain states are how the mind appears when viewed from respectively first and third–person perspectives, and that these aspects of mind are mutually irreducible. If so, the discovery of the neural

correlates of given experiences will not settle the fundamental differences amongst these theories. Meantime, it is evident that no ontological view is automatically privileged, phenomenal consciousness should not be defined in a way that *presupposes* the outcome of this debate. This practice however, is widespread in the scientific and philosophical literature. For M. Velmans for example, D. Dennett simply declares first–person access to phenomenal qualities to have no place in third–person science, and, therefore, no ultimate place in an understanding of consciousness at all. And that is of course is a nonsense. Why? Dennett considers the self as an entity that is spread across the brain both in time and space and includes both conscious and unconscious processes within the brain that are responsible for causing behavior. He is well known as the critic of „Cartesian theatre", an idea about the place „where it all (subjective experience) happens". Sometimes, as perhaps Dennett will put it, philosophers switch from „cartesian dualism" to less obscure „cartesian materialism" where the picture of a „ghost in the machine" has been replaced by the picture of a „brain in the machine". As Dennett writes, „those who claim to be materialists while still hanging onto the Cartesian Theatre with all its alluring imagery, he says are trapped in "Cartesian materialism", "the view that there is a crucial finish line or boundary somewhere in the brain, marking a place where the order of arrival equals the order of 'presentation' in experience" (Dennett, 1991, 107). Dennett's view on freedom and will (Dennett, 2003) enables to overcome the „old-fashioned" free will problem according to which free equals conscious states equals mental states as an ultimate cause of persons' behaviour. He has painted a picture of human freedom without any spirits moving us. Similarly, romantic love minus Cupid's arrow is stiil worth yearning for. So, could there be a consensus between researchers who take the existence of conscious phenomenology to be both self–evident and ontologically primary, with those who give no credence to that phenomenology at all? The answer to this question dependes on understanding what „giving

credence to phenomenology" actually means. If it means, for example to hold *a priori* an antireductive position in investigating consciousness, the answer will be No. But *what* are the grounds for holding an antireductive stance as a starting point, *before* the investigation even begins? Velmans, as the proponent of this idea is, as I think, wrong at this point and furthermore he moves back from and not towards a scientific understanding of consciousness, as he claims. One can only agree with Velmans in that „it is a mistake to *define* consciousness in a way that begs this question". At the same time insistence on a particular ontology, „*once a definition of "consciousness" is firmly grounded in its phenomenology, investigations of its ontology and its relationships to entities, events and processes that are not conscious can begin, and this may in time transmute the meaning (or sense) of the term*" (Velmans, 2009, 139) leads him to beg the question himself. Unsurprisingly, prevailing confusions and debates concerning the meaning of the concept of consciousness have raised skepticism towards solving the problem of the nature of conscious states. What is at stake is the common *perennial question* — What is the explanandum? People feel a need to justify their attitudes and actions. An „ideal model" is that of a rational, conscious, responsible and free agent. Scholars have demonstrated in a number of experiments several ways of separation of action from the experience of will. People feel conscious will in action they did not anticipate, e.g. in confabulation of intentions they revise what they think they intended to do after their action is complete. In „posthypnotic sugession" hypnotized subject accepts the instruction and often invents explanations following suggested actions. Patient with „allien–hand–syndrom" (Sacks, 1985) does not feel conscious will in anticipated action, he feels his hand as allien, performs an act for which he is responsible not being aware of it. In a „locked–in–syndrom" patient is conscious, he intends to perform an act, but is unable to act and suffers (Damasio, 1999). Everyday experience brings many examples of unconsciously caused action where people will revise their attitudes to

justify their action. If a person does something that is inconsistent with preexisting desire he feels uncomfortable and aims at avoiding thoughts in an dissonant relationship (cognitive dissonance). Classical cases of unconsciously performed actions have been studied in automatisms. As demonstrated, absence of experienced volitional act leads to attribution of causes and performance of events due to mysterious forces (e.g. spiritual seanses, Ouija boards, automatic writing, ghosts). The study and explanation of „ideomotor action" make paranormal „explanations" outrageous. Supposing that will is an experience rising from perceiving a causal link between thought and action it should be possible to lead people to experience willful action. This idea was the background for conducting an experiment known as „I Spy" study. Authors of the experiment were interested whether people will feel they willfuly performed an action that was performed by someone else. D. Wegner and S. Wheatley were inspired by the already mentioned household Quija board. Here goes a rough description of the experiment. Participant (P) and Confederate (C) P and C were asked to sit facing each other across a small table with a square board mounted atop a computer mouse. P and C were supposed to place their fingertips on side of the board in order to move the mouse together. They were asked to move the mouse in slow circles and to move a cursor around a computer screen. The screen showed about 50 small objects from the book *I Spy*. The experimenter explained that the study would investigate people's feelings of intention for acts, how feelings come and go (in time). P and C were instructed to stop moving the mouse every 30 seconds and rate each stop they made for personal intentionality. They made ratings on scales each of which consisted of line with endpoints „I intended to make the stop". P and C were also told they would here music and words through headphones during the experiment, P was lead to believe he heard different words than the C. The words served to prime thoughts about pictures on the screen. The C however heard just instructions to make movements at particular times. The P heard

the word consistent with the stop 30 seconds before, 5 seconds before, 1 second after the C stopped on the picture. The experiment has shown that timing of the thought in relation to the action plays a signifacnt role. When P was reminded of a picture on the screen 1 or 5 seconds before were forced to move the cursor they have reported having performed the movement willingly. Reminding them 30 seconds before the movement or 1 second after the forced movement, the sense of intentionlity decreased. According to the authors, experiment has proved that the experience of will can be created by manipulation of thought and action in accord with the following principles: a) priority principle, b) consistency principle and c) exclusivity principle. According to the first principle the thought should precede action at a proper interval. As the experiment has demonstrated, experience of will depends on the timely occurence of thought prior to action. Second principle is grounded in the idea that thought should be compatible with action and according to the third principle — thought should be the only apparent cause of action. The last principle suggests that people will be particularly sensitive to the possibility that there are other causes of an action besides their own thoughts. When their own thoughts do not appear to be the exclusive cause of their action they experience less conscious will. Finally, the „I Spy" study revealed the illusionary nature of mental causal chain — illusion of causal/time succession from, so to speak, the left to the right — from mind to behaviour. As demonstrated, the experience of consciously willing an action and the causation of the action by the person's conscious mind might become distinct. For Wegner the tendency to confuse them is the source of the illusion of conscious will. So, the experience of conscious will is an illusion in the sense that it is produced by the perception of apparent causal sequence relating one's conscious thought to one's actions. In reality, as he says, this may not be the causal mechanism at all. We have thoughts of what we will do, we can develop causal theories relating those thoughts to our actions. We come to think of these prior

thoughts as intentions, we feel that these intentions have causal force even though they are just previews of what we may do. That the sense of will is not directly connected to the causal mechanism reminds us of D. Dennett's (Dennett, 1991) taking an „intentional stance" towards people. It is based on viewing causation not in terms of underlying mechanism but rather of agents who have desires, intentions that cause their acts. And he seems to be right here. Differences in conceptualizing the problem of consciousness has led many philosophers towards identifying its characteristic as an illusion with nonexistent phenomenon. For S. Blackmore (2005) consciousness is truly a curious illusion. At the same time she is not saying that consciousness does not exist, as her critics claim: „I mean that consciousness is not what it seems to be". An illusion is *something* that is not what it seems to be". And for this reason Blackmore claims that the science of consciousness is built on false premises. D. Wegner contrary to S. Blackmore claims that the experience of conscious will is a feeling that helps us to appreciate our authorship of things we do. For him it is important to understand how conscious will might be an illusion and not that it is an illusion as a whole. Conscious will is an illusion in the sense that „*the experience of consciously willing an action is not a direct indication that the conscious thought has caused the action*" (Wegner, 2002, 2). M. Velmans argues that inspite of D. Wegners insights on how conscious experiences relate to brain processes it is necessary to outline in what sense conscious will is *not* an illusion. He agrees with Wegner's causal story of how conscious will arises and enters causally into subsequent mental processing. For Velmans, however, this causal story remains a first–person story. Conscious free will is not an illusion in the sense that this first–person story is compatible with and complementary to a third–person account of voluntary processing in the mind/brain. So, in what senses can conscious free will be considered as an illusion? Letś suppose that the causal role of any conscious experience in a conscious mental process can be said to be an illusion. Mental process then might be conscious (a) in the sense that one is conscious *of* it, (b) in the sense that it *results* in a conscious experience, and (c) in the sense that conscious experience plays a *causal role* in that process. According to Velmans, Wegner has shown that the experienced will is a *representation* of what is going on in the mind/brain. Preconscious decision making processes can be said to become conscious once they *result* in a conscious free will experience. According to Velmans, conscious will is illusionary when an experience of will can arise from voluntary processes and represent them without governing them. The illusion is based on the feeling that our conscious will determines our decisions and actions. But how can experience of will *arise* from „voluntary processes"? What preceeds what? Voluntary processes, as Velmans argues are not an illusion. Although conscious representations *of* those processes can be inaccurate, they can also be accurate and evolution has ensured that mental representations (conscious or not) are more often right than wrong. The nature of both voluntary processess and conscious representations has been left unexplained and therefore looks rather mysterious.

Suggested readings

Blackmore, S.: *A Very Short Introduction to Consciousness.* Oxford University Press, Oxford, 2005.

Libet, B.: Do We Have Free Will? In: *Journal of Consciousness Studies* 8–9 (1999) pp. 47–57.

Velmans, M.: Why conscious free will both is and isn't an illusion. In: *Behavioral and Brain Sciences,* volume 27, no5, october 2004, 649 — 659.

Wegner, D.: *The Illusion of Conscious Will.* Cambridge: MIT Press, 2002.

VI. On rational and emotional action

Key words: *reason, consciousness, impairment, passions, explanation*

Recent advances in neuroscience have dramatically improved our understanding of human emotional states. With the help of new technologies and models scholars begin to unfold the „mystery" of emotional life. Confusions in contemporary emotion studies are due to the traditional model of a person as a rational conscious agent. I intend to highlights two problematic aspects of this prevailing model, the relation between emotion and reason and between emotion and consciousness. Firstly, it is claimed that the difference between emotions and thoughts does not transcend their mutual interconnectivity. Secondly, the conscious content and emotional responses are both products of specialized emotion systems that operate unconsciously. Both claims are supported by experimental findings and clinical practise. Minds without emotions are like souls on ice — cold, lifeless creatures devoid of any desires, fears, sorrows, pains or pleasures. One can hardly disagree with these words of Joseph LeDoux, leading authority in neural science. Feelings, fear, joy or sadness are inseparable from our everyday experience. We seldom search for the „place" from where our emotions come from. They seem to appear suddenly, they can change slowly, in one moment we may feel happy and at a glance we can start to cry. So, it seems that emotions come and go „as they like" without our willful participation. So far so good. But it is sometimes not easy to explain what is and what is not consistent with our commonsense intuitions about ongoing emotional states. Ability to „read" minds of others, feel empathy or anger, predict behaviour and action of others, makes us quite successful folk psychologists.

However, a plausible theoretical approach in studying and explaining emotions requires more. Scientists, philosophers aim to give a systematic account of what emotions are, how they operate in the brain, and why they play such an important role in our lives. Paradoxicaly, flourishing cognitive science in the first half of twentieth century ignored emotions both in practice and theory. Cognitive scientists started to study cognitive states of mind independently of emotions. Research has been concentrated on solving primarily questions on how we think, solve problems, perceive, control our actions, remember a fact etc. After the decline of the behaviorist era it has become popular to approach mind as a function of a system. Fascinated by the metaphor of a computer cognitive scientists and philosophers created a new cognitive–computer picture of mind. Heterogeneous mental states were approached on the background of an interconnected input–information processing–output scheme. Functionalism in philosophy and psychology reined in studying cognition for decades. The artificial separation of emotions in cognitive science in the early days of cognitive revolution has made an important contribution to both experimental and theoretical inquiry and helped to establish a new approach to the mind. But, soon it became evident that human mind is a rather specificaly engineered *feeling* machine with a rich evolutionary history. Inspired by new facts from neuroscience, evolutionary theory, psychology and anthropology, scholars reconsidered their attitude to the proclaimed mysterious „inner world" and started to search for new levels of analysis, hypotheses and research programs. Problematic nature of emotions has been expressed in the following questions, for example: What is the relation between emotions, thoughts and will? Are emotions innate? What does it mean to have the capacity for self–control? Are we responsible for

our actions, volitions, decisions? Why is it important to unfold the neuronal mechanisms of emotional states? Do comatose patients feel? Where do feelings come from? What are conscious emotions for? As the contemporary state of field suggests, there is no single general approach or theory of emotions. Scientists and theoreticians are rather confronted with a collection of various psychological phenomena closely related to the problems of the nature of mind, free will, consciousness and overall behaviour. Complex and interdisciplinary nature of any proposed research on emotion thus represents a great challenge for scholars of various disciplines. In the following text I intend to point up two problematic aspects in emotion studies. Firstly, the relation between reason and emotion, secondly misunderstandings concerning emotions and the conscious/unconscious dichotomy. Confusion about the nature of emotion has a long tradition in philosophy, psychology and neuroscience. According to a traditional model of a person, the role of emotions in the overall picture of human behaviour was underestimated. To have control over one's actions pressuposess to act as a maximaly rational conscious agent with a minimum impact of emotions. Since the time of ancient Greeks theoreticians found it compelling to separate reason from passion, thinking from feeling, cognition from emotion. Passions were mostly considered as heritige of our animal predecesors. Plato placed emotions within his broad conception of mind based on the distinction between reasoning, desiring and feeling. For him passions, desires and fears are a kind of „enemies" for thinking. Emotions were analogous to wild horses that have to be reined in by the intellect which he thought of as a charioteer. For Plato this picture represented the struggle between passions and the intelect, rational part of the soul which was identified with wisdom. René Descartes analysed in detail emotions such as anger, happiness, fear, love, anxiety in his treatise *Passions of the Soul*. Passions represented „perceptions, feelings, emotions of the soul which we relate to it and which are caused, maintained and fortified by some movements of the spirits

(Descartes, 1967, 344). He characterized the struggle between body and soul as a war of the volitional soul against passions and bodily processes. The moral of Descartes' approach encouraged to use experience and reasom in such a way as to avoid zeal. Reason is suppossed to prevent the soul from being able wholly to control its passion. Knowledge of the nature and functions of passions is thus a necessary prerequisite in order to control passions and act rightly towards others. Conception based on firm and determinate judgements respecting the knowledge of good and evil in man's life had a fundamental influence on psychology and physiology of the present time. David Hume's philosophical tradition has been less generous to emotions which were treated as realms of animal and flesh. Immanuel Kant analysed the concept of soul (freedom, god) independently from experience, as an unempirical idea of the reason. A man is able to act moraly only when he is free, when his will is unconstrained and acts „within itself". Morality of an act is in Kant's conception closely related to the war against the sensational, „natural" in men. The role of the critic of practical reason is to avoid pretensions of the empiricaly determined reason to become a genuine determinative of the will. Kant's moral categorical imperative is considered as the activity of „pure will" and „pure reasom". By the end of nineteenth century Darwin, James and Freud had written on emotions extensively and given emotion a privileged place in the science of mind and behaviour. Charles Darwin in his work *The expression of the Emotions in Man and Animals* (1872) emphasized the unity of expressing emotions in humans independently of their culture and also the continuity of expressing emotions in animals and humans. He has observed that human facial expressions are sometimes homologous with those of primates. Darwin's identification of several expressions of emotions, happiness, joy, fear, sadness, anger, surprise, disgust as pan–cultural, is still recognized. Ethological tradition emphasized inheritance of complex behaviour patterns in humans and in the similar vein attributed an adaptive value to emotions. In his work famous

ethologist Konrad Lorenz revived darvinian ideas in contrast with the widespread behaviorist theory and practise. William James proposed that conscious emotion feeling is the perception of automatic nervous system changes caused by external stimuli via a reflec arc. His ideas influenced work of one of the leading contemporary neuroscientist Antonio Damasio who enriched theorizing on emotions with a new inspirative perspective. Data from neurobiological studies demonstrated a close connection between thinking and feeling. Relevancy of feeling within decision making has been often illustrated by the classical case of Pineas Cage (Damasio, 1994). Heavy brain damage to specific areas of both prefrontal lobes had an immense impact on the personality of Pineas. Gradualy, he has lost his capacity to decide, plan and to find a permanent job. His impairment led to a total emotional breakdown, social isolation and loss of personality. Treating thoughts and feelings as interconnected, surely, does not imply that they do not differ. For Joseph LeDoux conscious emotional feelings and conscious thoughts are in some sense similar. They both involve the symbolic representation in working memory of subsymbolic processes carried out by systems that work unconsciously. The difference between them is due to the fact, that a) emotional feelings and thoughts are generated by different subsymbolic systems, b) emotional feelings involve many more brain systems than thoughts (LeDoux, 1998, 299). Emotions create a flurry of activity devoted to one goal while thoughts do not. We can easily daydream while doing other things, reading, eating etc. But, when faced with danger or challenging emotional situations we do not have time, for example to kill, the whole self gets absorbed in the emotion itself. The word „emotion" does not refer, according to LeDoux, to something that the mind or the brain really has or does, it is more a convenient way of talking about aspects of the brain and its mind. Various classess of emotions are mediated by separate neural systems that have evolved for different reasons. There is no such thing as „emotion" faculty and neither a single brain system dedicated to it. Brain systems

that generate emotional behaviour are highly conserved through many levels of evolutionary history. All animals, including people have to satisfy certain conditions to survive in the world and fulfill their biological imperative to pass their genes on to their offspring. Insects, worms, fish, birds, rats and people need to obtain food and shelter, protect themselves from bodily harm and procreate. The neural organisation of particular behavioral systems (f.e. systems underlying fearful, sexual or feeding behaviours) is pretty similar accross species. Brains are not the same, but in order to understand what it means to be human involves an appreciation of the ways in which we are like other animals as well as the ways in which we are different. As emotional beings we think of emotions as conscious experiences. When we try to understand for example love or what a feeling is, why it occurs, where it comes from, the subjective feeling itself does not have much to do with it. We are not primarily interested in *what it is like for an individual to feel*, as some philosophers recently claim (Chalmers, 1996). Their effort is based on an idea according to which the „hard problem" in explaining human mind and consciousness lies in the very phenomenon of subjective feeling, in the way we *consciously* experience, feel our inner states and surroundings. However, experimental research, clinical practise, study of impairments of conscious mind suggest that generation of emotional responses does not require for the most part the presence of consciousness. Subjective emotional states like all states of consciousness are the result of information processing occuring unconsciously (LeDoux, 1998, 37). Damasio made it explicit when he wrote „emotions happen to us rather than things we will to occur" (Damasio, 1999, 19). That is also why he considers feeling feelings, identified with consciousness, as only the tip of the mental iceberg. But, once emotions occur they become powerful motivators of future behaviours. Besides being useful, they can also have pathological consequences. As a number of case studies show, mental problems reflect a breakdown of emotional order. In his approach on consciouness Damasio emphasizes unity of states

of consciousness, self and emotions. Emotions are associated with one type of the evolutionary older consciousness, namely *core consciousness*. Patients whose core consciousness is impaired do not reveal emotion by facial expression, body expression or vocalization. The entire range of their emotional life is usually missing. Another type of impairment shows that for example patients with preserved core consciousness but impaired extended consciousness have normal background and primary emotions. The fact that „biological machinery underlying emotion is not dependent on consciousness" (Damasio, 1999, 43) is best illustrated by a following case study. Due to an extensive damage to both temporal lobes (and hippocampus, amygdala) young man, David, cannot learn any new fact. Surprisingly, David seemed to manifest preferences and avoidances for certain persons even if he could not recognize any of them. Damasio put this idea to empirical test and designed a good–guy/bad–guy experiment. David has been engaged in three distict types of human interaction: a) extremely pleasant, b) emotionally neutral and c) bad mannered, boring. David's exposure to the good, to the bad and to the indifferent has been measured and compared. Then David was asked to look at sets of four photographs that included the face of one of the three individuals in the experiment, his task was to identify among them a friend. David has chosen a) over 80% of the time, b) chosen with a probability of a chance and c) bad guy was almost never chosen (in reality a pretty young women!). Experiment demonstrated that when core consciousness remains intact, nonconscious preference of patient's brain could generate actions commensurate with the emotional value (Damasio, 1999, 45). Experiment not only supported plausibility of dividing consciousness into more evolutionary evolved types but also contributed to understand the complexity of relations between emotions, conscious feelings and the brain activity. One may find himself/herself happy or nervous and yet be at a loss to why this or that particular state happens. Emotional feelings do not come from nowhere, their substrate is a set of neural

patterns in maps of selected structures. Animals similarly as humans, can solve a number of problems without being conscious of what, how and why they do so. Consciousness elevates thinking to a new level, but it isn't the same thing as thinking. Emotional feelings result when we become consciously aware that an emotion system of the brain is active. Transformation from emotion to conscious feeling is a complex journey, Damasio recognizes at least five steps from emotion to feeling to feeling of feeling ((Damasio, 1999, 283). When emotional systems function in an animal that also has the capacity for conscious awareness the conscious emotional feeelings occur. We know that we have an emotion when the sense of the feeling of self is created in our minds, until there is this self, there exist well–orchestrated responses, which constitute an emotion, brain representations which constitute a feeling. Problematic nature of the conscious/unconscious divide is closely connected with a popular but at the same time theoreticaly controversial phenomenon of pain. Pain is the consequence of a state of local dysfunction in a living tissue — tissue damage causes sensation of pain but also causes regulatory responses such as reflexes and may also induce emotions on its own (Damasio, 1999, 71). We can come to know that we have pain and that we are having an emotion associated with it, provided there is consciousness. When you picked up a hot plate and burned the skin of your fingers you had pain and might even suffered from having it. What has happened, can be put in neurobiological terms as following: The heat had activated a large number of thin and unmyelinated nerve fibers, C–fibers, available near the burn. Then, the heat destroyed several thousand skin cells and the destruction released a number of chemical substances in the area. Several classes of white blood cell concerned with repairing tissue damage were called to the area, a number of chemicals activated nerve fibres on their own, joining their signaling voices to that of the heat itself. Once the activation wave started in the nerve fibers it travelled to spinal cord and chain of signals was produced across several neurons and synapses. As the result

of the succession of signals neurons were temporaly activated which produces a neural pattern — the conditions to generate a sensation of pain had been met. Knowing that you have pain requires something *else* that occurs after the neural patterns that correspond to the substrate of pain (nociceptive signals) are displayed in the appropriate areas of the brain stem, thalamus, cerebral cortex and generate an image of pain, a feeling of pain. It is a brain process that interrelates neural patterns of tissue damage with the neural patterns that stand for you such that another neural pattern can arise — of you knowing which is just another name, according to Damasio for consciousness. If this latter mentioned process does not take place, you will never know that there was tissue damage in your organism. Vast number of emotional responses are generated automatically producing changes in facial expression and posture along with the changes in heart rate and control of blood circulation. For instance, many of such responses are present in comatose patients in whom consciousness is suspended. Evaluation of the state of their nervous system consists of establishing whether the patient reacts with facial and limb movements to unpleasant stimuli. Even if the importance of the above mentioned research for further diagnosis and therapy is evident, there still remains plenty of open ethical, legal and social questions (Illess, 2006). We can see that contrary to our everyday intuitions and also to traditional philosophical conceptions, pain and emotion are not the same thing. A case study of a patient in whom was dissociation between pain as such and emotion caused by pain makes it vivid. The patient was suffering from a severe case of refractory trigeminal neuralgia — *tic douloureux*. Under this condition the nerve that supplies signals for the face sensation in which even innocent stimuli, light touch of the skin, trigger an excruciating pain. After two days since the surgeons have operated on him Damasio visited the patient. He realized that the patient had become an entirely different person, relaxed, and when he asked him about the pain he replied: „the pains were the same, but I do not

feel them". So, the operation had abolished the emotional reactions that the sensory patterns of tissue dysfunction had been engendering. Suffering was gone. To claim that emotions represent a fundamental aspect of human life is a commonplace. Explaining and understanding their nature is not so straightforward. So far, experimental and theoretical research revealed new mechanisms and principles which condition the activity of numerous emotational states. Problems in attributing the proper place of emotions in our mental life and a search for an appropriate methodology has been reflected in a number of models, cognitive and noncognitive theories. In giving insight into how the most personal aspects of our mind work scientific understanding of emotions could help us understand what may go wrong when this part of our mental life breaks down. Research in neuroscience, neuropsychology and cognitive science together with novel technologies brought new insight also in investigating and understanding the relation between reason and emotions. As recent studies show, emotions play an important part in practical decision making, a strict dichotomy between „feeling heart" a „cold reason" has been definitely abandoned. The existence of an entity called „pure reason" is also an illusion, because a majority of goal formating, problem solving, planning includes a variety of cognitive–emotional features. An immense presence and role of unconscious states in human decision, planning and acting undermined the traditional ideas on conscious states. Most of traditional approaches and conceptions of emotions (reason, free will, consciousness) thus have been challenged. Human mind or emotional states are no more considered as immaterial „ghosts in a machine" or a threat for rationality. The lasting struggle between thought and emotion may be probably soon resolved by a harmonious integration of reason and passion in the brain. Experimental research and clinical practise have demonstrated a need in reconsidering the model of a rational conscious agent. It is evident that free action cannot be identified with conscious action, will is far from being an inner unobjectifiable mental

„entity“. Understanding motives, causes and effects of action presuposess to take into account man's inclination towards irational decisions, misunderstandings, and a number of cognitive biases grounded in underestimation or overestimation the capacities of mind. Theoreticans can no longer „weap under the carpet“ new empirical and theoretical findings. In reality they may be helpful in clarifying the meaning and use of terms, questioning superficial dichotomies and asking new questions. Disolving the terminological fog which covers concepts of consciousness, will, rationality, conscious/unconsious processes, volitional acts would play an important role in overcoming conceptual misunderstandings and strengthening explanatory power of recent theories.

Suggested readings

Damasio, A. *Descartes's error: Emotion, Reason and the Human Brain.* New York: Grosset/Putnam, 1994.

Damasio, A. *The Feeling of What Happens.* London: William Heinemann, 1999.

Darwin, Ch. *The expression of the emotions in man and animals.* Chicago: University of Chicago Press, 1965/1872.

VII. Impairments of consciousness

Key words: *personality change, bipolar disorder, depression, schizophrenia, sleeping disorder*

Phineas Gage (1823–1860) is one of the earliest documented cases of severe brain injury. Gage is the index case of an individual who suffered major personality changes after brain trauma. As such, he is a legend in the annals of neurology, which is largely based on the study of brain–damaged patients. Gage was foreman of a crew of railroad construction workers who were excavating rocks to make way for the railroad track. This involved drilling holes deep into the rock and filling them with dynamite. A fuse was then inserted, and the entrance to the hole plugged with sand, so that the force of the explosion would be directed into the boulder. This was done with a crow bar–like tool called a tamping iron. On 13th September, 1848, 25 year old Gage and his crew were working on the Rutland and Burlington Railroad near Cavendish in Vermont. Gage was preparing for an explosion by compacting a bore with explosive powder using a tamping iron. While he was doing this, a spark from the tamping iron ignited the powder, causing the iron to be propelled at high speed straight through his skull. It entered under the left cheek bone and exited through the top of the head, and was later recovered some 30 yards from the site of the accident. John Martin Harlow, the doctor who attended to him, later noted that the tamping iron was found “several rods behind him, where it was afterward picked up by his men smeared with blood and brain”. The

tamping iron was 3 ft. 8 inches in length and 1.25 inches in diameter at one end, not 1.25 inches in circumference, as reported in the newspaper report on the left. It tapered at one end, over a distance of about 1 ft., to a blunt end 0.25 inches in diameter, and weighed more than 6 kg. Whether or not Gage lost consciousness is not known, but, remarkably, he was conscious and able to walk within minutes of the accident. He was then seated in an oxcart, on which he was transported three–quarters of a mile to the boarding house where he was staying. Here, he was attended to by Harlow, the local physician. At the boarding house, Harlow cleaned Gage's wounds by removing small fragments of bone, and replaced some of the larger skull fragments that remained attached but had been displaced by the tamping iron. He then closed the larger wound at the top of Gauge's head with adhesive straps, and covered the opening with a wet compress. Gage's wounds were not treated surgically, but were instead left open to drain into the dressings. Within a few days of his accident, Gage's exposed brain became infected with a "fungus", and he lapsed into a semi–comatose state. His family prepared a coffin for him, but Gage recovered. Two weeks after the accident, Harlow released 8 fluid ounces of pus from an abscess under Gage's scalp, which would otherwise have leaked into the brain, with fatal consequences. By 1st January 1849, Gage was leading an apparently normal life. His contractors, who regarded him as the most efficient and capable foreman in their employ previous to his injury, considered the change in his mind so marked that they could not give him his place again. He is fitful, irreverent, indulging at times in the grossest profanity, manifesting but little deference for his fellows, impatient of restraint of advice when it conflicts with his desires, at times pertinaciously obstinent, yet capricious and vacillating, devising many plans of future operation, which are no sooner arranged than they are abandoned in turn for others appearing more feasible. In this regard, his mind was radically changed, so decidedly that his friends and acquaintances said he was "no longer Gage." The damage to Gage's frontal cortex had

resulted in a complete loss of social inhibitions, which often led to inappropriate behaviour. In effect, the tamping iron had performed a frontal lobotomy on Gage, but the exact nature of the damage incurred to his brain has been a subject of debate ever since the accident occurred. This is because the damage can only be inferred from the path of the tamping iron through Gage's skull, which in turn can only be inferred from the damage to the skull. Gage's skull was damaged in three places: there is a small wound under the left zygomatic arch (cheek bone) where the tamping iron entered; another is located in the orbital bone in the base of the skull behind the orbit of the eye; and the third, and largest, wound is in the top of the skull, where the tamping iron exited. The exit wound was enormous, and never healed. It can be seen today in Gage's as an irregularly–shaped triangular hole, about 2 inches wide and 4 inches in circumference, and another, nearly 3 inches in circumference. These are separated by one of the flaps of skull that was replaced by Harlow upon arriving at Gage's boarding house. Because the circumference of the wound in the frontal bone is much larger than the maximum diameter of the tamping iron, it is difficult to determine precisely the trajectory of the iron and where it exited Gage's skull. In 1994, Hannah Damasio and her colleagues at the University of Iowa used neuroimaging techniques to reconstruct Gage's skull. The conclusion of this study was that Gage incurred damage to both the left and right prefrontal cortices. But according to computer–generated three–dimensional reconstructions of a thin slice computed tomography scan of Gage's skull, the damage to Gage's brain was limited to the left hemisphere. Nevertheless, the case of Phineas Gage made important contributions to early modern neurology. So, what of Phineas Gage himself? Unable to return to his previous job as a foreman after his accident, Gage is said to have travelled around New England, and even to Europe, with his tamping iron trying to earn money. It is also said that he even displayed himself as a curiosity at Barnum's Museum in New York. However, the story of Phineas Gage is as much folklore as it is fact. Not only

the exact nature of the neurological damage Gage sustained, but also the details of his life after the accident, are disputed to this day. It is known that, from 1851 until just before his death, Gage worked as a coach driver, first in a livery stable at the Dartmouth Inn, in Hanover, New Hampshire, for about 18 months, and then in Chile for some 7 years. At some point in 1859, with his health deteriorating, Gage went to live with his mother. He died in San Francisco on 20th May, 1860, some 12 years after his accident, of complications arising as a result of epileptic convulsions. *Bipolar disorder* (previously termed „manic–depressive illness“) is a relatively common and chronic psychiatric condition in which patients experience episodes of mania and depression, usually with intervening periods of relative mood stability. Bipolar disorder is associated with cognitive and behavioural difficulties and in severe cases psychosis can present in both the manic and depressive states. Often beginning in adolescence or early adulthood, bipolar disorder has a profound negative effect on interpersonal, social, family and vocational outcomes and is a risk factor for substance abuse and suicide. While the exact cause of bipolar disorder has not been elucidated, there are likely to be multiple contributors to the patho–aetiology of the disorder. A number of studies have implicated several areas of the brain and have focused attention on abnormalities in the intra–cellular processes of brain function, such as cell receptors and neurotransmitter effects. Most recently, studies have explored the possibility of neural degeneration as a potential common final pathway in the disorder. A number of pharmacological treatments have been shown to be effective for the treatment of the manic and depressive states of the disorder and in the prophylaxis of episodes. Common agents used include lithium, anti–convulsants and anti–psychotics (mood stabilizing action). Other non–pharmacological treatments, such as electroconvulsive therapy, and different psychotherapeutic approaches, e.g. social rhythms therapy, are also effective and can be preventative or even life saving in certain cases. *Schizophrenia* is a serious psychiatric disorder

that causes people to lose touch with reality. The schizophrenic disorders are characterised by profound disruption of thinking and perception that affects the most fundamental human attributes including cognition, language, perception and sense of self. Schizophrenia is one of a group of related mental conditions called psychoses. These disorders may be severe and chronic and all have very different patterns of onset and outcome. Intelligence and consciousness are not always affected by these disorders, although certain cognitive impairments may develop over time. Perception may become disturbed, the individual may become extremely perplexed, thinking may be vague and obscure, and his/her expression in speech often becomes incomprehensible. *Sleep disorders* disturb the normal sleep pattern and may cause fatigue, irritability, anxiety, and impaired memory and concentration. Some sleep disorders are also associated with increased morbidity and mortality. According to the DSM–IV and ICD–10 classification systems, the primary sleep disorders can be divided into two main groups: dyssomnias and parasomnias. Dyssomnias are disorders in which the length, timing or type of sleep are different from normal. Some sleep disorders are found in conjunction with psychiatric disorders, e.g. depression, neurological disorders, e.g. Parkinson‘s disease and other medical disorders e.g. endocrine problems. *Panic disorder* is a type of anxiety disorder and is characterised by recurrent episodes of panic attacks and the development of fear, worries and anxiety regarding the possibility of future attacks, along with several physical symptoms. The unpredictable recurrence of new panic attacks may cause significant fear and lead to severe avoidance behaviour and anticipatory anxiety. Panic attacks are characterised by sudden and unexpected distinct periods of intense fear, nervousness or apprehension, terror, panic or discomfort. They are often accompanied by physical symptoms, such as: shortness of breath, dizziness, palpitations, angina, excessive perspiration, trembling, nausea or abdominal distress; and cognitive symptoms such as: depersonalisation or derealisation, and the fear of losing

control, going crazy, having a heart attack, or even dying. These somatic experiences lead many patients to seek care in non–psychiatric settings, such as emergency rooms and physicians' offices and patients suffering from PD use healthcare services at rates higher than usual. In one study, PD was found to be associated with greater demand for general, emergency and psychiatric services than any other psychiatric diagnosis. In the search for physical causes to explain the physical symptoms associated with panic disorder, the diagnosis is often missed. It can take up to eight years before the diagnosis is made, despite regular contact with medical services.

Suggested readings

Damasio, A.: *Descartes's error: Emotion, Reason and the Human Brain.* New York: Grosset/Putnam, 1994.

Gray, J.: Consci*ousness: Creeping up on the Hard Problem.* Oxford University Press: Oxford, 2004.

Revonsuo, A.: *Consciousness: The Science of Subjectivity.* New York: Psychology press, 2010.

VIII. Idols of the reason

Key words: *idols, cognitive biases, racionality, illusions, misrepresentation*

In everyday experience, perceiving the surrounding reality, man experiences a number of illusions. Whether conscious or unconscious these illusions form an insearable part of our daily lives. We can briefly distinguish outer and inner illusions, former are known as *optical illusions* and later, less recognized, inner illusions. Both types of illusions or *idols* (as John Locke will put it) are a great source of „data" in studying and explaining human mind. It has been argued here that cognitive science neglects an important source of insight into the human mind: the effects created by manipulating minds. Over the centuries magicians have learned how to perform acts that are perceived as defying the laws of nature, and that induce a strong sense of wonder. This paper argues that the time has come to examine the scientific bases behind such phenomena, and to create a science of magic linked to relevant areas of cognitive science. Concrete examples are taken from three areas of magic: the ability to control attention, to distort perception, and to influence choice. It is shown how such knowledge can help develop new tools and suggest new avenues of research into human perception and cognition. Imagine a ball tossed into the air that suddenly disappears. Or someone uncannily predicting exactly what you will do in the next few minutes. These fantastical scenarios exist not only in science fiction, but are experienced by anyone

who has ever witnessed a skilful conjurer in action. Over the centuries magicians have learned how to perform acts that are perceived as defying the laws of physics and logic, leaving an audience baffled and amazed. Yet there is nothing otherworldly about these effects — they are created entirely by natural means. We argue here that there is great scientific potential in studying the ways that most people can be made to believe in such "impossible" events, even if only for a few seconds. In particular, we argue that the effects by magicians can provide us with valuable tools to investigate human perception and cognition. Although a few attempts were made in the distant past to draw links between magic and human cognition, this knowledge has been largely neglected by modern psychology. We propose that the time has come to examine these phenomena more closely, and to connect them to current theories and methodologies for exploring the human mind. The history of science has shown that theories often stem from knowledge obtained from practical applications, for example, thermodynamics from the development of steam engines. We argue that a similar situation exists here: Over the centuries, magicians have accumulated considerable knowledge about inducing strking effects in human observers. We believe that this knowledge can be systematized and used as a source of insight into mechanisms central to human perception and cognition. In addition, these effects also suggest new methodological techniques to investigate the relevant processes. We will illustrate these points by examining three general methods used by magicians: misdirection, illusion, and forcing. There is a common belief that magicians hide their techniques, or methods by relying on speed. But it is simply false that "the hand is quicker than the eye": most manipulations are carried out at a normal pace. Rather than relying on speed, the success of an effect (i.e. the experience of the spectator) usually relies on *misdirection*, the diversion of attention away from its method, so that the audience does not notice how it was produced. This reliance on misdirection to achieve "invisibility" is closely related to recent findings in vision

science that only a small part of the information that enters our eyes, the part that is attended, enters our conscious awareness. Magicians have known this for centuries, and have accumulated considerable practical knowledge about how to control the relevant mechanisms. They have proposed a framework that distinguishes between physical misdirection, based on the physical properties of the stimulus, and psychological misdirection, based on control of higher–level expectations. Physical misdirection refers to the control of attention via stimulus properties; this is similar to the concept of exogenous control found in psychology, in which certain stimulus properties automatically capture our attention. The goal is to create areas of high interest that capture the spectator's attention, while the method is covertly carried out in an area of low interest. A wide range of techniques have been found to be effective. For example, an important rule in magic states that the audience will look where the magician is looking. This has an interesting connection to recent work showing that eye gaze leads to automatic shifts of visual attention (). Stimulus properties such as movement, high contrast, and novelty are also regarded as important; this also has been found in recent empirical studies. Although many such cues have already been investigated scientifically, the magician's use of them suggests that they will have considerably more power when combined correctly. Many methods involve attentional capture, in which attention is pulled away by an irrelevant task. These could be used to improve our understanding of how capture operates. For example, psychologists so far have focused on properties that capture attention in space, paying less attention to issues of time. Magicians have found that control can also be achieved through repetition, or "off beat" moments, which lead to a momentary relaxation (such as after a joke), during which the spectator's attentional "hold" is relatively weak. Magicians also use non–verbal signs such as body posture to manipulate the level of vigilance, which then affects attentional allocation. Experiments based on this form of attentional control could provide valuable

insights into attentional modulation over time. *Psychological misdirection* controls spectators' attention by manipulating their expectations; this is similar to the concept of endogenous control found in psychology, in which attentional orienting is determined by a person's goals and intentions. The magician's aim is to reduce suspicion that a deceptive method has been used. For example, he may require a secret prop that needs to be gotten rid of by putting it back into his pocket. If the action of putting his hand into his pocket seems normal and/or justified (e.g., he put his hand into his pocket on previous occasions), the action will cause far less suspicion and will therefore be far more likely to go unnoticed. Another way of reducing suspicion is by keeping the audience in suspense as to what they are about to see. As long as the spectators don't know what to expect they will not know which aspects of the routine are important, and so will be unlikely to direct their attention to those aspects needed for the effect. Related to this, a key rule in magic states that magic tricks should never be repeated. Indeed, it has been shown that both repetition and prior knowledge about what the spectator will see increases the likelihood that the observer will detect the method. Psychological misdirection can also be done via the false solution, which a magician will highlight so as to divert attention from the real solution. For example, a magician can pretend to have been caught out, so that the spectator will ignore all other less obvious solutions. Once the spectator has been sent down this garden path this false solution can be revealed to be false. However, by this time, most of the tracks have been covered and he will find it difficult to discover the correct solution. This is likely related to the Einstellungs effect, the finding that once an idea comes to mind, alternatives are often not considered. Work in vision science has shown that much of vision is essentially a form of intelligent hallucination. To perceive depth, for example, the visual system must recover the third dimension from the two–dimensional image available on the retina. However, since multiple solutions are usually possible for a given image, the result must be

obtained by applying assumptions of some kind. This approach, however, can sometimes lead to errors, which take the form of illusions. Two types of illusions are typically employed by magicians: optical, which involve physical factors, and cognitive, which involve psychological ones. Many conjuring tricks, especially those of the stage illusionist, involve optical illusions, which rely on tricks such as intricate mirror combinations and perspectives. For example, by manipulating the perspective of an object the true size of a box can be distorted, leaving plenty of room to hide an elephant. Other techniques, such as Pepper's ghost illusion, use mirrors and special lighting to make an object appear and disappear in full view of the audience; this effect can also be used to make one object seem to morph into another. Some of these illusions could be implemented as the basis for new forms of investigations into visual perception. There is another „group" of illusions which are highly relevant in understanding human mind and behaviour — called *cognitive illusions.* Most sleight of hand magicians tend to rely on "higher level" cognitive factors, rather than the "smoke and mirrors" used by the stage illusionist. An example of this is the "vanishing coin illusion". Here, the spectator perceives the magician transferring a coin from one hand to the other, with the coin then vanishing. But in reality, the coin never changes hands, it is instead secretly concealed in the hand and so remains out of sight. The key to sleight of hand involves discovering the extent to which the "false" action can be altered to make the spectators still feel they are seeing the "real thing". Interestingly, spectators often report having seen a "real" event, even though it never took place. Why might such effects occur? The finite speed of neural transmission causes a delay of approximately 100 ms between stimulus arrival and conscious percep. One way of compensating for this is to "predict the present", predict the outcome of an event before it has been completely processed. This strategy is particularly useful in situations that require rapid response, such as skilled driving or sports. But such predictions can also make us vulnerable to

deception. Effects such as the vanishing coin illusion and the vanishing ball illusion are experienced whenever the available evidence is consistent with the prediction made by the spectator. Effects of this kind may serve as useful starting points for the empirical nvestigation of the subjective aspect of perception. Imagine picking a card from a deck of playing cards. To your astonishment, you find that the magician has predicted your choice. Although you felt like your choice was free, in reality it was highly controlled. The process by which your choice can be systematically influenced is known as forcing. This has interesting connections to recent work showing that observers often confabulate about the reasons for their choices and finally their acts.

Suggested readings

Damasio, A.: *The Feeling of What Happens.* London: William Heinemann, 1999.

D. Kahneman, P. Slovic, A. Tversky: *Judgment under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press, 1982.

F. Mikš: *Gombrich: Tajemství obrazu a jazyk umění.* Brno: Barrister a Principal, 2008.

Bibliography

Arendtová, H.: Řeč a metafora. In: *Myšlení o divadle II.*, M. Petříček (ed.). Praha: Hermann, 1993, 51–78.

Aristoteles: *Poetika.* Praha: Svoboda, 1996.

Bacon, R.: *Nové Organon.* Praha: Svoboda, 1990.

Black, M.: More about metaphor. In: *Metaphor and thought,* A. Ortony (ed.). Cambridge: Cambridge University Press, 1996, 19–41.

Blackmore, S.: *Consciousness: An Introduction.* London: Oxford University Press, 2011.

Block, N., Flanagan, O, Guzeldere, G.: *The Nature of Consciousness: Philosophical and Scientific Debates.* Cambridge: MIT Press, 1996.

Block, N.: "On a confusion about a function of consciousness". In: *Behavioral and Brain Sciences*, 1995, 18(2), pp. 227–272.

Chalmers, D.: 'Facing up to the problem of consciousness'.In: *Journal of Consciousness Studies*, 1995, 2(3), pp. 200–219.

Crane, T.: Concepts in Perception. In: *Analysis* 48 (June), 1988, 150–53.

Crick, F. *The Astonishing Hypothesis: The Scientific Search for the Soul.* New York: Scribner, 1994.

Chalmers, D. *The Conscious Mind.* New York: Oxford University Press, 1996.

Damasio, A. *The Feeling of What Happens.* London: William Heinemann, 1999.

Damasio, A. *Descartes's error: Emotion, Reason and the Human Brain.* New York: Grosset/Putnam, 1994.

Darwin, Ch. *The expression of the emotions in man and animals.* Chicago: University of Chicago Press, 1965/1872.

Dennett, D.: *Consciousness Explained.* Boston: Little, Brown, 1991.

Dennett, D.: *Freedom Evolves.* London: Allen Lane, 2003.

Davidson, D.: Mentálne udalosti. In: Čin, myseľ, jazyk. Bratislava: Archa, 1997.

Descartes, R. *The Philosophical Works of Descartes.* Vol.I. Cambridge: Cambridge University Press, 1967.

Descartes, R.: *Vášně duše.* Praha: Mladá fronta, 2002.

Evans, G.: *The Varieties of Reference.* Oxford: Oxford University Press, 1982.

Eysenk, M., Keane, M.: *Kognitivní psychologie*. Praha: Academia, 2008.
Gazanniga, M., Ivry, R., Mangun, G.: *Cognitive Neuroscience: The Biology of the Mind*. New York: W.W. Norton, 2002.
Sacks, O.: *Muž, který si pletl manželku s kloboukem*. Praha, Mladá fronta, 1993.
Gáliková, S.: *Psyché*. Bratislava: Veda, 2007.
Gáliková, S., Gál, E.: *Antológia filozofie mysle*. Bratislava, Kalligram, 2003.
Churchland, P. S.: *Brain–Wise: Studies in Neurophilosophy*. Cambridge: MIT Press, 2002.
Illess, J. *Neuroethics*. Oxford: Oxford University Press, 2006.
James, W. *Principles of Psychology*. New York: Holt, 1890.
LeDoux, J. *The Emotional Brain*. New York: Touchstone, 1998.
Churchland, P. M.: *The Engine of Reason, The Seat of the Soul: A Philosophical Journey into the Brain*. Cambridge: MIT Press, 1995.
Gray, J.: Consci*ousness: Creeping up on the Hard Problem*. Oxford University Press: Oxford, 2004.
Libet, B. Do We Have Free Will? In: *Journal of Consciousness Studies*, 6, No. 8–9, 47–57, 1999.
Kahneman, D., Slovic, P., Tversky, A.: *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press, 1982.
„Epiphenomenal and Supervenient Causation", *Midwest Studies in Philosophy*, Vol. IX, Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein, eds. Minneapolis: University of Minnesota Press, 1984, 257–70.
Kim, J.: „What is ‚Naturalized Epistemology'?", *Philosophical Perspectives*, Vol. 2, 1988, 381–405.
Kim, J.: *Mind in a Physical World*. Cambridge: MIT Press, 1998.
Mikš, F.: *Gombrich: Tajemství obrazu a jazyk umění*. Brno: Barrister a Principal, 2008.
Popper, K.: *Hľadanie lepšieho sveta*. Bratislava: Archa, 1995.
Putnam, H.: *Mind, Language and Reality. Philosophical Papers, vol. 2*. Cambridge: Cambridge University Press, 1975.
Revonsuo, A.: *Consciousness: The Science of Subjectivity*. New York: Psychology press, 2010.
Ryle, G.: *The Concept of Mind*. London: Hutchinson, 1949.
Sacks, O.: *The Man Who Mistook His Wife for a Hat, and Other Clinical Tales*. Summit Books, 1985.
Sacks, O.: *Na čem si stojím* ... Praha: Dybbuk, 2009.
Searle, J. (1992): *The Rediscovery of Mind*. Cambridge: MIT Press.
Velmans, M.: How to define consciousness — and how not to define consciousness. In: *Journal of Consciousness Studies*, 16(5), 2009, pp. 139–156.
Wegner, D.: *The Illusion of Conscious Will*. Cambridge: MIT Press, 2002.
Wegner, D., Weathley, T.: Apparent mental causation: Sources of the experience of will. *American Psychologist*, 1999, 54: pp. 480–491.

Wittgenstein, L.: *Tractatus logico–philosophicus* (z nem. prel. P. Balko, R. Maco). Bratislava: Kalligram, 2003.
Wittgenstein, L.: *Modrá a Hnedá kniha* (z angl. prel. D. Kamhal). Bratislava: Kalligram, 2002.
Wittgenstein, L.: *Philosophical Investigations*. Oxford: Basil Blackwell, 1953.

Prof. PhDr. Silvia Gáliková, CSc.

Introduction to the Philosophy of Mind

Firs edition

Peer reviewers
Doc. PhDr. Ján Rybár, Ph.D.
Doc. Matúš Porubjak, Ph.D.

Proofreading PhDr. Blanka Holigová
Graphic © Ladislav Tkáčik

fftu

Publisher
Faculty of Philosophy and Arts · Tranava University in Trnava
Hornopotočná 23 · 918 43 Trnava
filozofia@truni.sk · Http://fff.truni.sk

in cooperation with

Towarzystwo Słowaków w Polsce
ul. św. Filipa 7 · 31–150 Kraków
zg@tsp.org.pl · Http://www.tsp.org.pl

ISBN 978–83–7490–611–1