Fenk, A. (2012). Free will, the mind-brain problem, and testability. In: M. G. Weiss & H. Greif (eds.) Ethics – Society – Politics. Papers of the 35th International Wittgenstein Symposium, Vol. XX. Kirchberg am Wechsel: Austrian Ludwig Wittgenstein Society, pp. 75 – 77

Free Will, the Mind-Brain Problem, and Testability

August Fenk

Klagenfurt, Austria | august.fenk@uni-klu.ac.at

1 Free will and Libet's results

Libet et al. (1983: 623) investigated the readiness-potential "that precedes a freely voluntary, fully endogenous motor act" and found that the "onset of the cerebral activity clearly preceded by at least several hundred milliseconds the reported time of conscious intention to act." This reported time was "the subject's recall [...] of his initial awareness of wanting or intending to move." The authors conclude that cerebral initiation can begin "before there is any (at least recallable) subjective awareness that a 'decision' to act has already been initiated cerebrally." (Maybe a somewhat careless wording: Who could ever be aware of the fact that his or her intention has already been initiated cerebrally?)

These results and interpretations unleashed a controversial discussion about the impact on ethics: Do they question free will, one of the basics of our self-concepts and of moral responsibility? A maybe overshooting reaction in view of the fact that the lags reported rely on subjective memories; also note the authors' qualification "(at least recallable)". It seems, moreover, that readiness- (or motor-) potentials – cf. the terms "expectancy wave" and "Contingent Negative Variation (CNV)" – reflect cortical processes being more generally associated with goal-directed cognitive activities such as problem solving (Fenk 1978) or sentence comprehension (Kutas/Hillyard 1980). Do Libet's measures really guarantee an isolation of components specifically relevant for the initiation of "a freely voluntary, fully endogenous motor act"? Is it, after all, possible to follow the instruction to produce, now and then and "without any preplanning" (Libet et al. 1983: 625), one or two previously defined movements? And could such prescribed movements really be classified as "fully endogenous"?

If, however, in Libet's experiments and in affirmative studies of other laboratories (e.g. Soon et al. 2008) all measures were appropriate and all confounding variables – such as a possible bias in the subjects' recall – eliminated, they would enforce a rethinking: Free will is an illusion, decisions are taken in the brain. But I will show in the next section that they would enforce such a rethinking only within a specific, though very common mind-body theory.

Brembs (2011: 930) emphasizes that the "historical use of the term 'free will' has been inextricably linked with one variant or another of dualism" – positions that have, like the "metaphysical" (in the sense of "untestable"?) concept of *free will*, "fallen into irrelevance". *Free will* sensu Brembs shows itself in an evolutionary advantageous unpredictability of an animal's actions in "always the same settings" (ibid.: 935) such as in pursuit evasion. Apart from the well-known "sameness-problem": Could such an unpredictability reflect something relevant for those concepts of *free will* dominating the philosophical debate, at least as a prerequisite? Instead of such questions I shall discuss (in Section 2) which variant of dualism would admit the traditional understanding of *free will* as a conscious agent and prerequisite for moral responsibility. Interactionism (physical⇔mental) will result as the only possibility. Sections 3 and 4 investigate whether interactionism can, in the light of epistemic virtues such as testability and parsimony, compete with monism, or with epiphenomenalism (physical⇒mental), where free will is illusory from the onset.

2 The assumption of free will in different mind-body theories

Figure 1, left panel, classifies the positions compared in this paper. But is this an exhaustive classification? Popper (1996: 5) for instance describes himself as a "Cartesian *dualist*", and since he explicitly assumes an "*interaction* between *physical* and *mental* states", we can localize his position in our box called "interactionism". But he declares himself, moreover, as a *pluralist* who accepts "the reality of a *third world*, [...] roughly, the world of the *products* of our human minds." Figure 1 does without a further box for "pluralism", because the body-mind problem concerns, also for Popper, the relations between the physical (world 1) and the mental (world 2). The dependency between the physical, including physiological processes, and the psychic – a term that I here use synonymously with the "mental" or the "conscious" – is bidirectional in interactionism and unidirectional in epiphenomenalism.

Genealogical relations would deserve a separate diagram and more space for discussion. For instance: Was (Leibniz's classical) parallelism superseded by monism, and (e.g. Feigl's 1963) monism by functionalism? Or is functionalism only a variety of monism (Bechtel 2010)? Is it still a mind-body theory or rather a structure-function theory? Does Roth's (1987) constructivist approach supersede or only paraphrase (e.g. Rohracher's 1953) epiphenomenalism? I cannot see much difference except Roth's emphasis on the perceptual world as our brains' construct (ibid,; 235); thus my brain that I can watch in a brain-imaging experiment could not be identical with the real brain that produces my mental image (ibid.: 238f).[1]

In his seminal study Feigl (1963) modifies his earlier "double-language theory", i.e., the mind-body problem as the problem of a mentalist language that cannot be reduced to that language we use to describe "macro-behavior" and physiological processes ("micro-behavior"). It would be wiser, says Feigl (ibid.: 447), "to speak instead of *twofold access* or *double knowledge*. The identification, I have emphasized, is to be *empirically* justified, and hence there can be no *logical* equivalence between the concepts (or statements) in the two languages."

Carnap (1963: 886) criticizes "Feigl's reference to certain facts as 'evidence' for the identity view" and suggests formulating the question of a justification of that view "in the metalanguage [...] as a question concerning the choice of a language form. Although we prefer a different language, we must admit that a dualistic language can be constructed and used without coming into conflict with either the laws of logic or with empirically known facts."

Apart from empirical testability, Feigl (1963: 474) claims a further advantage of his solution: It "differs quite funda-

[1] This example recalls Feigl's (1963: 473f) "autocerebroscopic" *gedankenexperiment*. But Feigl is not mentioned in Roth, nor is any proponent of epiphenomenalism mentioned in Feigl and Roth, respectively.

mentally from materialistic epiphenomenalism in that (1) it is *monistic,* whereas epiphenomenalism is a form of dualistic parallelism; (2) the 'physical' [is not] the primary kind of existence, to which the mental is appended as a causally inefficacious luxury, or 'shadowy' secondary kind of existence."

This criticism of epiphenomenalism is meant at e.g. Rohracher's (1953) *Theorie der letzten Wirkung:* The mental is conceived as the ultimate effect of the physical (Figure 1, right panel); the assumption of any retroaction of the mental on the physical is "superfluous" (ibid.: 159) and free will a delusion. Rohracher would rebut the qualification of his theory as dualistic "parallelism", because he understands parallelism exactly as the theory that *denies* any action of the physical on the mental and vice versa. And he explicitly rebuts the qualification "materialistic" (ibid.: 154) because of his view of the mental as something fundamentally different from the underlying processes.[2]

To sum up: In epiphenomenalism all mental processes are caused by neural activities; thus it is from the beginning compatible with Libet's results and incompatible with free will. Parallelism in the usual sense also excludes any influence of the mental on the physical, and thus also any influence of free will. And what could influence what in monism if there is no distinction between different sets of events? Interactionism seems to be the only theory that provides the possibility of free will. But can it cope with epistemological criteria?

3 Mind-body theories in the light of epistemic virtues

How could the physical produce/influence the mental? And how could the mental initiate/control the physical? These questions do not arise in parallelism or monism. And where they arise – the first one in epiphenomenalism, both in interactionism – they remain unanswered. But all of these theories make statements concerning the possibility of (mutual) dependencies. Which of them hold in the light of epistemic criteria such as testability, compatibility with "common sense", and Occam's principle of parsimony? In Hume's problem Occam's razor and common sense favor the very same dissolution (Fenk 2010); in the mind-body problem however they apparently favor different positions.

Ad *testability*: Feigl claims that his monism is empirically justifiable, i.e., in principle testable. But he drastically weakens that claim through his notice "that the identity thesis is a matter of epistemological and semantic interpretation, and does not differ in empirical consequences from a carefully formulated parallelism." (ibid.: 472) Maybe he was, when writing that paragraph, not fully aware of the fact that parallelism, in whatever form, is dualistic. In his comment on Feigl, Carnap rather maintains the linguistic analytical approach and remains extremely vague concerning the possibility of relevant empirical findings: Like Feigl he believes "that the evidence available today provides good reasons for the assumption that this [monistic] language will also function well in the future." (Carnap 1963: 886) But what are these good reasons? And would a lag of the mental (Libet et al. 1983; Soon et al. 2008) be compatible with monism and with a "carefully formulated" parallelism?

Everyone knows that the mind can be influenced e.g. by drugs – obviously not directly, but via body. And for those actively working in neuroscience it is not really surprising that some of the relevant (sub)cortical processes precede the respective psychical changes. They can study the effects of (experimentally manipulated) neural activities on behavior and on (reports about) mental processes. But never the other way round: How could the "mental" be observed, or experimentally manipulated, without affecting neural processes? In "brain reading" and its applications, which raise "ethical issues concerning the privacy" of thought (Haynes and Rees 2006:523), mental states can be reconstructed only through our knowledge of their correlation with patterns of neural activities. Such correlations as well as effects of the physical on the mental are the only testable thing and the only thing considered as given in epiphenomenalism, while interactionism also claims retroactions that are not testable. But if epiphenomenalism explicitly excludes the possibility of such retroactions, this is a 0-hypothesis and as such again not really testable (Fenk 2010:85), and moreover a 0-hypothesis that disperses a naturalist concept which allows any event to be both, cause and effect.

Concerning *common sense*, at least in the common sense of that term, interactionism is the clear winner. We are convinced that our sensations are evoked by stimuli from inside and outside our body, but that we also can choose between alternatives – e.g. where to look and thus also which stimuli to affect our visual system. Our language regime reflects and reinforces that we experience ourselves as "decision making" and "voluntarily acting", and it presupposes even involuntary effects of the mental on the physical, such as "psychogenic" factors contributing to "psychosomatic" disorders. From such a perspective, all other mind-body theories must appear rather artificial. These are either incompatible with effects of the mental on the physical, as in epiphenomenalism, or with the idea of effects in general: Parallelism denies such effects in any direction from the very beginning. And which components could act on other components in monism (see below) if there are only two different "ways of knowing the same event"?

The principle of *parsimony*, however, is only realized in epiphenomenalism: It accounts for the actual empirical knowledge with a minimum of existential assumptions. Parallelism pays a high prize for its parsimony regarding mind-body interactions, namely the need to explain the mysterious power that has established synchrony – but see Libet's results – between the correlated events. Monism also claims parsimony for itself: Feigl (1963) directly addresses Ockham's razor (ibid.: 386) or principle of parsimony when e.g. deleting Spinoza's "third substance" (ibid.: 449), and indirectly when replacing "the duality of two sets of correlated events [...] by the less puzzling duality of two ways of knowing the same event – one direct, one indirect" (ibid.: 473). But this new duality is, to my mind, not an attempt to a parsimonious explanation but to avoid anything that would deserve an explanation. And it is again puzzling: Is this "same event" a reintroduction of what he criticizes in Spinoza as the "third substance" or "reality-in-itself" – in order to complete his saying (ibid.: 447) concerning a "twofold access" (to what?) or "double knowledge" (of what?)?

4 Concluding remarks

Brembs (2011: 935) argues that freedom could be dissociated from both, consciousness and will. Regarding the latter he cites John Locke: "I think the question is not proper,

[2] In Searle (1994: 3) consciousness is also caused by brain processes, but this consciousness "is not some extra substance or entity" – as in logical behaviourism, where "mental properties *are* physical properties" (Marek 1994: 142).

whether the will be free, but whether a man be free". Such attributions would not be obviated by the results of the neurophysiology of action (cf. Trivers 2011: 55), but would shift some of our problems to the organism as a whole and would raise some new questions concerning determinism: "Empirical" questions, as claimed in Balaguer (2009: 20)? And how to prove once and for all the claims of partial indeterminacy (Balaguer) and unpredictability (Brembs), i.e., the *absence* of regularity in specific behavioral domains?

Free will as a conscious agent however makes sense only in interactionism. But the assumption of retroactions of the mental on the physical is neither testable nor is it necessary for the analysis of behavioral and mental processes. Thus, the epistemic principle of parsimony favors the less popular epiphenomenalism. An explicit exclusion of the possibility of such retroactions would, on the other hand, amount to an again not testable 0-hypothesis. And if epiphenomenalism is *not* explicit in that point, then it only recalls what kind of answers are within or beyond the reach of neuropsychology. Such a position would, however, correspond with Wittgenstein's understanding of "common sense", with his last dictum in the Tractatus and with his estimation of possible outcomes of philosophical investigations (Wittgenstein 2006): To aim at "*complete* clarity /.../ simply means that the philosophical problems should *completely* disappear."(§133; see also §119 and 124)

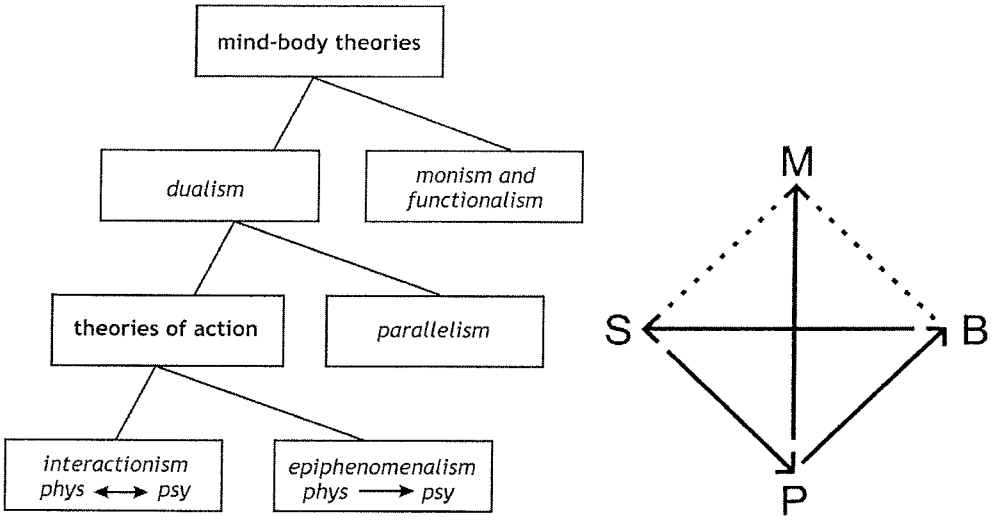

Figure 1: Classifies the most influential psychophysical theories (left panel) and illustrates epiphenomenalism (right panel): Physiological(P) processes not only control behavior(B) – which in turn modifies stimulus-patterns(S) – but also cause the mental(M) events; these do not interact with or retroact upon neural activities or anything else.

References

Balaguer, Mark 2009 "Why there are no Good Arguments for any Interesting Version of Determinism", *Synthese* 168, 1-21.

Bechtel, William 2010 "How Can Philosophy be a True Cognitive Science Discipline?", *Topics in Cognitive Science* 2, 356-366.

Brembs, Björn 2011 "Towards a Scientific Concept of Free Will as a Biological Trait: Spontaneous Actions and Decision-Making in Invertebrates", *Proc. R. Soc.* B 278, 930-939.

Carnap, Rudolf 1963 "Herbert Feigl on Physicalism", in: Paul Schilpp (ed.), *The Philosophy of Rudolf Carnap*, Lasalle, IL: Open Court, 882-886.

Feigl, Herbert 1963[1958] "The 'Mental' and the 'Physical'", in: Herbert Feigl, Michael Scriven, and Grover Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, Vol. II: *Concepts, Theories, and the Mind-Body Problem*, 370-497.

Fenk, August 1978 "Langsame Hirnpotentiale bei verschieden strukturierten Aufgaben", *Zeitschrift für Psychologie* 186, 89-112.

— 2010 "Hume's Problem, Occam's Razor, and Wittgenstein's Standpoint of Common Sense", in: Elisabeth Nemeth, Richard Heinrich, and Wolfram Pichler (eds.), *Image and Imaging in Philosophy, Science, and the Arts,* Kirchberg am Wechsel: The Austrian Ludwig Wittgenstein Society, 84-86.

Haynes, John-Dylan and Rees, Geraint 2006 "Decoding Mental States from Brain Activity in Humans", *Nature Reviews Neuroscience* 7, 523-524. Retrieved from www.scopus.com (21.11.2011).

Kutas, Marta and Hillyard, Steven 1980 "Event-Related Brain Potentials to Semantically Inappropriate and Surprisingly Large Words", *Biological Psychology* 11, 99-116.

Libet, Benjamin et al. 1983 "Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity", *Brain* 106, 623-642.

Marek, Johann 1994 "On the Relation of the Mental and the Physical", in: Roberto Casati (ed.), *Philosophy and the Cognitive Sciences: Proceedings of the 16th Int. Wittgenstein Symposium*, Vienna: Hölder-Pichler-Tempsky, 39-145.

Popper, Karl 1996 *Knowledge and the Body-Mind Problem*, London/ New York: Routledge.

Rohracher, Hubert 1953 *Die Arbeitsweise des Gehirns und die psychischen Vorgänge*, München: Johann Ambrosius Barth.

Roth, Gerhard 1987 "Erkenntnis und Realität: Das reale Gehirn und seine Wirklichkeit", in: Siegfried Schmidt (ed.), *Der Diskurs des Radikalen Konstruktivismus*, Frankfurt: Suhrkamp, 229-255.

Searle, John 1994 "The Problem of Consciousness", in: Roberto Casati (ed.), *Philosophy and the Cognitive Sciences: Proceedings of the 16th Int. Wittgenstein Symposium*, Vienna: Hölder-Pichler-Tempsky, 1-10.

Soon, Chun Siong et al. 2008 "Unconscious Determinants of Free Decisions in the Human Brain", *Nature Neuroscience* 11/5, 543-545.

Trivers, Robert 2011 *Deceit and Self-Deception*, London: Allen Lane.

Wittgenstein, Ludwig ^{3}2006 [1953] *Philosophical Investigations*, Malden (MA): Blackwell Publishing.