Chapter

30 Trialogue: commentaries on "Are mental illnesses disorders of consciousness?"

Comments on Panksepp and on Vogeley & Newen

G. Lynn Stephens and George Graham

In a contribution to an earlier collection on the neurobiology of mental illness, George Heninger (1999, p. 89) remarks as follows: "At an idealistic theoretical level it would greatly simplify matters if all mental disorders were as straightforward as phenylketonuria." Why does Heninger say that? Because, he writes, "here there is a specific biochemical pathogenesis" (p. 89). However, we hasten to note that phenylketonuria, assuming that our Consciousness Thesis is correct, really is not a mental illness. Although phenylketonuria is a syndrome with a mental deficiency, it is not a disorder in and of conscious experience. So, it fails to qualify as a mental illness. Heninger seems to admit as much, noting that its best descriptions are in neurology and not psychiatry texts (p. 91).

We are encouraged that each of our distinguished co-contributors to this section of the current volume seems to agree with us on that specific score. Mental illnesses occur in and of conscious experience. What makes them mental as opposed to non-mental is not their non-physicality (no room for mind-body dualism here), but the experiential influences that course through them.

Perhaps Panksepp is correct that one type of conscious experience in particular – namely, a loss of certain positive and basic emotional experiences of sorts shared by human beings with some species of non-human animals – leads to "mental strife of psychiatric significance". Certainly, this hypothesis seems a promising partial explanation of what happens, for example, in various anxiety disorders. Human beings, just as certain species of non-human animals, possess an evolved tendency to respond fearfully to some situations (the presence of snakes, for example), but not to others (the presence of trees, for example). Call these tendencies, as they are sometimes called in the literature, lurking fears. Our emotional equilibrium is disrupted by lurking fears and some individuals (perhaps those prone to anxiety attacks) may be oversensitive to the perceived presence of relevantly fearful stimuli in their environment.

We are dubious (is Panksepp also?) whether neuroanatomically locating which "brain states . . . have been evolutionarily prepared to be responsive" to the world in fearful ways will help us to decide between competing accounts of the origins of anxiety disorders. We doubt whether a mere anatomical hypothesis will identify which of two causal neurobiological routes is responsible for an anxiety attack; a "leftover" evolutionary one in terms of the continued salience of ancient stimuli or one in terms of classically conditioned stimuli. Competition between two such hypotheses must consider behavioral data as well as the learning histories of individual organisms (see Coltheart, 2006 for related discussion). Even if we locate the neural base of lurking fears in particular areas or systems of the brain, location alone will tell us little if anything about how those areas represent, encode and instantiate the processes responsible for anxiety attacks. Does an area help to produce attacks by encoding ancient response patterns, or by implementing mechanisms of classical conditioning that have the power to transform any current stimulus into one that elicits oversensitive fear?

The Neuropsychology of Mental Illness, ed. Stephen J. Wood, Nicholas B. Allen and Christos Pantelis. Published by Cambridge University Press. © Cambridge University Press 2009.

Kai Vogeley and Albert Newen (hereafter "V&N") claim that "the ability to distinguish one's own mental states from those of others" is central to self-consciousness, i.e. to "awareness of one's mental states as one's own" and to "the capacity to adequately ascribe mental states to others in order to explain and predict their behavior." Clearly, understanding how we acquire and exercise this ability represents an important part of our understanding of normal cognitive development. It may also contribute to our understanding of mental disorders. V&N draw attention particularly to autism and to certain phenomena associated with schizophrenia, such as thought insertion, experiences of alien control and verbal hallucinations.

We agree with V&N that both autism and delusions such as thought-insertion involve a failure to "differentiate between one's own and other's mental states." We want to point out, however, that a patient's problems with "self–other differentiation" are quite different in the two sorts of cases. These differences indicate the scope and complexity of problems that must be faced in trying to understand how people distinguish their own mental states from those of others, or may fail to do so in cases of mental disorder. We have written extensively about such problems elsewhere (see, for examples, Graham, 2004; Stephens & Graham, 2000). Here we wish to further elucidate their relevance to V&N's helpful discussion.

First, consider autism. According to a now widely (though not universally) accepted account, autistic subjects (and normal 3-year-olds) perform as they do on false belief tests because they lack a "theory of mind." In the case of autistic subjects, the lack is due to impairment and not expressive of a normal stage of conceptual development. Autistic children suffer from a deficient understanding of mentality. (Of course, "autism" is an extremely complicated and varied diagnosis, and no such generalization covers all cases. However, we assume that the just-mentioned attribution of impairment is sound for present illustrative purposes.) Uta Frith (1989), for example, explains that children suffering from autism do not "appreciate the difference between their own beliefs and someone else's beliefs, and that there can be different beliefs about a single event" (p. 159). They fail, she says, "to realize fully what it means to have a mind and to think, know, believe, and feel differently from other people" (p. 173).

Now contrast with thought insertion. Thought insertion also represents a striking breakdown in a subject's ability to differentiate his own thoughts from the purported thoughts of others. However, thought insertion represents a very different sort of problem of self–other differentiation from the one found in autism.

Normally, introspection enables persons not just to identify both the content of their thoughts or attitudes and their attitude type, but also their own self as the thinker or believer. I believe that *the car looks red*. I *believe* that the car looks red. *I* believe that the car looks red. In thought insertion, however, the subject finds specific episodes of thinking or believing occurring in his stream of consciousness which, judging from his introspective reports, he somehow fails to recognize as his own. Indeed, he takes them to be somebody else's thoughts.

Unlike young victims of autism on Uta Frith's (1989) account, persons suffering from delusions of thought insertion are not cognitively "blind" to the existence of minds *qua* minds: either their own or those of other persons. They recognize and employ the distinction between their own thoughts and those of others, even if, with regard to certain thoughts, they draw the distinction differently. Nor are they oblivious to the possibility that their own thoughts and those of others may differ in content. They have learned the general distinction between my thoughts and other peoples' thoughts, although in certain and often striking instances, they apparently mistake their own thoughts for someone else's thoughts. These thoughts appear to them to be alien.

It is not obvious just what precise mistake in attribution the subject makes in thought insertion. The subject locates the alien thoughts in his own mind or stream of consciousness, as opposed to believing, for example, that he is directly aware of or eavesdropping upon someone else's stream. However, since the thoughts he claims to occur in him (in his mind) he also claims to be someone else's, he seems to be reporting something that is self-contradictory. How can one and the same mental episode be both mine (in me) and not mine (someone else's)? Is there a coherent interpretation of the subject's delusion?

V&N offer a distinction that helps with this interpretative problem. V&N speak of two different ways of ascribing mental states to oneself, one of which involves an experience of agency and the other of ownership. In our own work we draw a similar

distinction between what we describe as an experience of agency and an experience of subjectivity. (We prefer to speak of subjectivity rather than of ownership because, in part, the language of ownership may connote a form of personal appropriation, of "owning up" to or taking responsibility for an episode. No such acceptance of responsibility occurs on the subjectivity side of the experiential ledger in thought insertion.) These distinctions help to remove the apparent contradiction in a person's claim that someone else's thoughts occur in his mind. They also suggest, as we have argued elsewhere, that there is a psychological link or similarity between thought insertion and various specific symptoms of schizophrenia, including delusions of control and of voices or verbal auditory hallucinations (see Stephens & Graham, 2000).

The main idea goes something like this: I may acknowledge (in the case of my body) that a fist clenching counts as my activity, in the sense that it happens to my body rather than someone else's body, but then deny that I myself clenched the fist (i.e. that I did it deliberately or voluntarily). Just so, I may say that the thought "Kill God" occurs in my mind (I am the subject), but deny that I am the author of or agent behind the thought. "Kill God" is not something that I, as a mental agent, *think*.

V&N mention a hypothesis of Christopher Frith which states that persons who experience their own activities, mental or bodily, as those of another agent or as due to an alien agency do so because they suffer from a breakdown in subpersonal cognitive systems that monitor connections between intentions and resulting activities. However, in our view, something other or more than monitoring failure is involved in the disorder. Although inserted thoughts seem to the person to be disconnected from his mental economy, they do not have the random character normally associated with transient, non-voluntary thoughts running through one's head. They seem purposive or characteristic of a mind, but not just one's own mind, and it is because of this appearance of being purposive that they are attributed to or experienced as those of another agent.

Ours may not be the right or remotely complete explanation of the experience of alienation (i.e. the experience of one's own thoughts as those of another), but clearly some explanation is needed. Just as clearly confusion about one's own mind and how it differs from another's lies at the heart of the conscious experience of a number of different mental disorders.

References

Coltheart, M. (2006). What has functional neuroimaging told us about the mind (so far)? *Cortex*, **42**, 323–331.

Heninger, G. (1999). Special challenges in the investigation of the neurobiology of mental illness. In D. Charley *et al.* (Eds.), *Neurobiology of Mental Illness* (pp. 89–99). Oxford: Oxford University Press.

Frith, U. (1989). *Autism: Explaining the Enigma*. Oxford: Blackwell.

Graham, G. (2004). Self-ascription: thought insertion. In J. Radden (Ed.), *The Philosophy of Psychiatry* (pp. 89–105). Oxford: Oxford University Press.

Stephens, G. L. & Graham, G. (2000). *When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts*. Cambridge, MA: MIT Press.

Affective consciousness and the psychiatric comfort zones of experienced life

Jaak Panksepp

Consciousness is such a multidimensional topic that many overlapping perspectives need to be considered for scientific illumination of this core problem of modern neuroscience. I found the contributions by Stephens & Graham (S&G) and Vogeley & Newen (V&N) to be synergistic with my own perspectives. The vigorous advocacy of S&G for a full acceptance of the phenomenology of conscious experience within psychiatric practice and our research endeavors is essential for dealing with human troubles in humane ways, and such visions can pave the road for the discovery of new, more subtle mind medicines through neuroscientific investigations. V&N provide a highly resolved cognitive view of consciousness with which I agree, but I encourage them to incorporate affective experience more explicitly into their analysis. I suspect that most cognitive abilities ultimately arose, in mammalian brain evolution, to service affective needs. Affect may be the most ancient form of consciousness, since it seems directly related to essential biological needs. Along with a few others (e.g. Denton, 2006), I have argued that the decoding of the affective strata of mind is an essential gateway for progress on the cognitive processes that may have evolved to optimize organismic search for the affective "comfort

zones" that support life. Of course, in humans cognitive activity has achieved a seeming self-sufficiency not evident in other species.

My own goal has been to understand the evolutionary underbelly of cognitive consciousness that is surely deeply affective – so deep, that some are prone to envision it as part of the dynamic unconscious. In line with Freud's perspectives, I think affects are fully experienced, but rarely talked about. V&N advocated conceptualizing consciousness in more cognitive terms than would I, even though the critically important aspect of cognitive views is the recognition of how deeply enmeshed our emotional processes are with our thoughts, as well as environmental events (i.e. situated cognitions). Of course, the critical scientific question is how we craft epistemological strategies for unraveling how experience is actually created within the brain, especially in a way that can positively impact psychiatric therapeutics. I suspect it will be rather easier to generate a clinically productive neuroscience of affect than of cognition. Much of what we have to say at the cognitive level will already have been said at one time or another, but the *detailed* neural analysis of higher cognitive functions is incredibly more difficult than that of lower affective functions, because the animal models are less robust. Although many who do not pursue detailed functional neuroscience have little motivation to distinguish between cognitive and affective processes of the mind–brain, since they are completely interpenetrant in intact organisms, the aim of science is to dissect the complexity of nature. There are many credible ways to distinguish affects and cognitions (Ciompi & Panksepp, 2004). Such distinctions provide potentially critically important considerations for clinically productive neuroscientific analysis (Panksepp, 2003, 2006).

For instance, I think the lack of ownership of experience that is often seen in schizophrenia, highlights how cognitive and affective processes become dissociated in this disconnection syndrome. In schizophrenia, higher and lower brain functions no longer operate as a coherent whole. Thus, the functional dissection of affect from cognition may be an essential scientific stepping stone for any evolutionarily coherent and psychiatrically relevant science of consciousness. Evidence abundantly indicates that psychiatrically relevant imbalances in consciousness reside heavily at affective levels, even though it is amplified and modulated by all sorts of ruminative activities. Affective imbalance may most commonly be the first-order symptoms of schizophrenia and other psychiatric disorders, while cognitive changes are second-order symptoms. Thus, I was a bit concerned that my esteemed colleagues focused such modest attention on affective consciousness.

My commentary arises largely from one pragmatic consideration: what can we scientifically achieve that is truly lasting and important, at the present time? I do feel abundant neuroscientific payoffs would emerge if we focused much more effort on primary-process forms of emotional-affective consciousness, the main *forces* for mental disequilibrium, than on the more visible cognitive (information-processing) dimensions of mind. Without affective turmoil, cognitions alone would rarely fall into the kinds of disequilibrium that lie at the core of psychiatrically significant problems in living, and the faulty object-relations they foster. However, I would qualify this claim: the sensory and homeostatic affects (e.g. as explicated by Denton, 2006), are surely of tertiary importance as compared with imbalances in raw emotional feelings – i.e. less important than cognitive factors.

Still, from the emotional vantage, the analysis of both V&N and S&G, although well within the mainstream of modern consciousness studies, may need to become immersed in the emotional forces that drive cognitive disequilibrium. Parenthetically I would add that for me a disciplined definition of "cognitive" is the permutation of information harvested by the external senses. Affects are largely intrinsic, within-brain evaluative processes, where information-processing metaphors no longer do much work. For instance, I would not define "consciousness as the subjective experience of one's own cognitive processes" as do V&N. There needs to be a clear distinction between raw primary-process experience and one's cognitive reflections on those experiences. I think we must begin to think in terms of evolutionary layering of consciousness, raising the possibility that more fundamental forms of consciousness may often become inhibited by the emergence of higher forms, and that extreme primary-process variants begin to re-emerge in sustained psychiatric distress.

Although cognitive reflections and affective feelings are highly interactive in many clinically pregnant ways, perhaps various primary emotional processes can serve as *endophenotypes* that can promote breakthrough thinking in psychiatric thinking and practice (Panksepp, 2006). Each meaningful psychiatric syndrome surely has a distinct set of critically important

emotional state processes that have become imbalanced, and it will be interesting to study how cognitions become enmeshed in those poorly regulated emotional *energies*. Indeed, if we readjust the imbalanced affective states, often it will be much easier to deal with cognitive disequilibrium. As noted by Kraemer (1993), psychiatric difficulties often resolve simply from pharmacological re-establishment of affective homeostasis with little need for cognitive intervention.

Of course, there is a downside to the powerful biological interventions that are now available. The discovery of hundreds of molecules to modify affect has now helped create a biological psychiatry without any clear vision of the mental apparatus. This is not good, even though it served as a corrective to the first half of the 20th century, where we had complex psychoanalytic ideas about mental processes, with no equally sophisticated neuroscience to bolster those ideas. It can be hoped that the goal of 21st century psychiatry will be to restore a balanced program, where the mind is brought fully back into neuroscientific views. Clear conceptions of how the brain generates core affective processes and the resulting cognitive entanglements must be a substantial part of that agenda.

Still, I abundantly agree with S&G's and V&N's emphasis of the importance of social-consciousness and cognitive-dynamics in psychiatric practice. The interface between human beings is both deeply affective and profoundly cognitive. Thus, the power of intersubjectivity has been one of the essential tools of all forms of psychotherapy, where the qualities of the therapist are commonly more important than the formal (manualized) qualities of the treatment. Presumably, at the foundation of all intersubjectivity there is a fairly small set of raw affective feelings that are critical for the quality of interaction, ranging from whether one is feeling lonely, insecure and feeling the pain of isolation or whether one is immersed in feelings of security, warm acceptance and trust. I would suggest such feelings are critically dependent on the degree of activity in the separation-distress (PANIC) networks of the brain, systems that are especially vigorously controlled by endogenous opioid, oxytocin and prolactin circuits. When such systems are satisfied (in homeostatic balance), then one can work more effectively with sexual feeling, nurturant tendencies, and certainly the degree of playfulness that can fill the intersubjective field.

Consider the PLAY circuits of the brain, barely recognized in psychiatric practice, that provide endless opportunities for enjoyable interactions that can be soul-healing (assuming, of course, that the "soul" is completely biological). I would go so far as to say that a child therapist who is not able to actually engage in rough-and-tumble playfulness with a young child, as opposed to simply partaking in toy- and game-facilitated interactions, is not using the full power of play therapy. Child psychiatrists who learn to put their toys to the side, and engage in real physical play, will open a very wide therapeutic door to all other interventions. The use of abundant rough-and-tumble play, each and every day, is much underutilized in developing children where we would like to abort the emergence of ADHD type symptoms (Panksepp *et al.*, 2003). Indeed, a case can be made that much of the symptomatic (but not sustained) therapeutic effect of psychostimulants in ADHD children is due to the robust anti-playful effects of such agents (Panksepp *et al.*, 2002a).

Also, most adult therapeutic environments would improve if a certain degree of the ineffable lightness of being, which can emerge from playful attitudes would more abundantly permeate the therapeutic interactions. There is a great deal of evidence suggesting that affective consciousness tends to be reciprocally related to higher cognitive activities – namely that lower limbic affectivity may not get fully aroused during abundant neocortical activity (Liotti & Panksepp, 2004). Obviously, thinking can sustain emotions as one dwells specifically on their troubles, but equally importantly, cognitive activity can also inhibit emotionality. Words can often get in the way. Thus, the uncovering and full acceptance of emotional feelings is as useful as any other aspect of consciousness within productive psychiatric interactions. We can anticipate that in the future there will be new medications that can work at the level of individual emotions, and some, perhaps oxytocinergics, may even help therapists to stay more effectively in the living moment.

In this context it is critically important to recognize the medial frontal cortical participation in feelings such as social cohesiveness, highlighted in V&N's important focus on the resting "self" work in brain imaging. Equally important to emphasize is that these brain regions participate in self-referential processing of all kinds of external perceptions (Northoff *et al.*, 2006), allowing lower and higher brain functions to

be blended. However, it is puzzling that a great deal of modern thinking, based on meager evidence, continues to assume that affect emerges fairly high in the neuroaxis even up to neocortical levels. Damasio's advocacy of a somatic-marker hypothesis is perhaps the most prominent of many examples of trying to place affect within the "read-out" processes of higher regions of the brain. Those regions, especially anterior medial and insular cortices, surely contribute much to how we experience life, but they may be incapable of affectivity without the lower sub-neocortical substrates we share with the other animals.

S&G also emphasize the role of experienced life in psychiatric disorders and changes in mental life that accompany addictions. I agree with their analysis. It allows us to better see how much of drug addiction is an attempt to self-medicate so as to improve affective homeostasis, and it is worth considering exactly what kinds of feelings certain addictive drugs promote. For instance, it is pretty clear that opioids can alleviate loneliness and replace the need for social relations (Panksepp, 1981). On the other hand, psychostimulants amplify euphoric engagement with the world, and promote states of social dominance (Panksepp *et al.*, 2002b).

Overall, I think all of us agree that neurophilosophically informed views of mental life, where lived subjective experience is never marginalized, provide better opportunities for major advances in psychiatric knowledge and practice than any form of ruthless "never-mind" reductionism that denies causal efficacy to the felt qualities of brain activities. A sophisticated and well-targeted consciousness view, where the positive social feelings of mutually experienced intersubjectivity is recognized as part of the healing equation, will allow us to use much lower doses of mind-modulating drugs. Indeed, we may then begin to look for new and more precise and gentle mind medicines, such as neuropeptide modulators, that may control very specific affective states (Panksepp & Harro, 2004). It is possible that such medicines, as well as the older ones that operate on more generalized brain-state regulatory mechanisms (e.g. the biogenic amine whole-brain "spritzers" that revolutionized psychiatry half a century ago), can be used at much lower doses in optimal psychotherapeutic environments. After all, psychotherapy can change brain dynamics and chemistries in richer, more symphonic ways than will ever be achieved with drugs.

It does seem that I am enamored by a more primitive (i.e. evolutionarily ancient) level of neurophenomenological analysis than either S&G or V&N. This comes from my own understanding of how the brain is organized in evolutionary layers, with the lower layers providing fundamental homeostatic-affective substrates for the experienced life, while the higher levels provide ever-increasing cognitive resolution to the opportunities and dangers of the world. Both levels of analysis are essential for any comprehensive picture, but perhaps the lower levels are currently much more susceptible to a rigorous neuroscientific analysis, if for no other reason than we have robust animal models where underlying causal issues can be studied in some detail. Information-processing approaches to mind tend to neglect that underbelly of mental life. Overall, affective neuroscience may advance psychiatric practice and understanding more than cognition-based views of consciousness.

References

Ciompi, L. & Panksepp, J. (2004). Energetic effects of emotions on cognitions – complementary psychobiological and psychosocial finding. In R. Ellis & N. Newton (Eds.), *Consciousness and Emotions* (Vol. 1, pp. 23–55). Amsterdam: John Benjamins.

Denton, D. (2006). *The Primordial Emotions: The Dawning of Consciousness*. New York, NY: Oxford University Press.

Kraemer, P. D. (1993). *Listening to Prozac*. New York, NY: Viking.

Liotti, M. & Panksepp, J. (2004). On the neural nature of human emotions and implications for biological psychiatry. In J. Panksepp (Ed.), *Textbook of Biological Psychiatry* (pp. 33–74). New York, NY: Wiley.

Northoff, G., Henzel, A., de Greck, M. *et al.* (2006). Self-referential processing in our brain – a meta-analysis of imaging studies of the self. *Neuroimage*, **31**, 440–445.

Panksepp, J. (1981). Brain opioids: a neurochemical substrate for narcotic and social dependence. In S. Cooper (Ed.), *Progress in Theory in Psychopharmacology* (pp. 149–175). London: Academic Press.

Panksepp, J. (2003). At the interface between the affective, behavioral and cognitive neurosciences: decoding the emotional feelings of the brain. *Brain and Cognition*, **52**, 4–14.

Panksepp, J. (2006). Emotional endophenotypes in evolutionary psychiatry. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, **30**, 774–784.

Panksepp, J. & Harro, J. (2004). The future of neuropeptides in biological psychiatry and emotional psychopharmacology: goals and strategies. In J. Panksepp (Ed.), *Textbook of Biological Psychiatry* (pp. 627–660). New York, NY: Wiley.

Panksepp, J., Burgdorf, J., Gordon, N. & Turner, C. (2002a). Treatment of ADHD with methylphenidate may sensitize brain substrates of desire. *Consciousness and Emotion*, **3**, 7–19.

Panksepp, J., Burgdorf, J., Gordon, N. & Turner, C. (2003). Modeling ADHD-type arousal with unilateral frontal cortex damage in rats and beneficial effects of play therapy. *Brain and Cognition*, **52**, 97–105.

Panksepp, J., Knutson, B. & Burgdorf, J. (2002b). The role of emotional brain systems in addictions: a neuro-evolutionary perspective. *Addiction*, **97**, 459–469.

The definition and the constitution of mental disorders and the role of neural dysfunctions

Kai Vogeley and Albert Newen

Comments

The two contributions by Panksepp and Stephens & Graham (S&G) provide important insights into the nature of mental diseases and its relation to consciousness. They focus on the role of emotional affects and pre-reflective processes for the constitution of mental diseases (Panksepp) and the relation between consciousness, mental diseases and their neural implementation (S&G).

Panksepp focuses on the significant scientific progress that has been made with respect to the nature of emotional affects and its neural implementation. As neuroscience is not yet able to read out any conscious experience individually, such as emotional awareness and self-awareness, Panksepp recommends that we should focus instead on raw emotional experiences that are assumed to be common to all human beings and probably even across mammalian species. His central claim with respect to mental disorders is that a major cause of mental disorders, at least with respect to affective disorders, can be traced back to a disturbance of raw emotional experience.

He proposes in a cross-species approach that psychological processes comprising basic or raw affective feelings arise from endogenous activity of neuronal assemblies. These have been evolutionarily prepared to become responsive to the world in certain behaviorally basic and significant, "instinctual" ways that evolved within a certain environment. Panksepp emphasizes the role of the non-conceptual dimension of consciousness for mental disorders that constitutes "primary-process consciousness." These raw feelings mediate emotional and motivational values and are essential for any higher-order awareness and intentionality. They are assumed to be a property of subneocortical brain processes that are considered a trans-species universal in the representation of raw feelings and which constitute a representation of a viscerosomatic homunculus or "the core self." Its localization in the paramedian area of the midbrain and diencephalon might well be correlated with the visceral and proprioceptive or interoceptive phenomena that are associated with emotions. As a pragmatic argument, Panksepp puts forward that psychopharmacological drugs such as antidepressants influence systems that are related to early and evolutionarily old processes. An ancient level of mentation might have thus been the foundation of higher-order consciousness that evolutionarily evolved or emerged from those simple and primitive forms of mental states. Subsequently, raw feelings can be transposed into emotional awareness of the reflexive type after being read out by the neocortex.

Panksepp emphasizes, as we do, that there is a basic type of consciousness that is a non-conceptual (or pre-reflexive) form of consciousness, and may include primitive "intentions in action" independent from any propositional attitudes. In greater detail we argue in favor of a distinction of five levels of self-consciousness including the non-conceptual self-acquaintance (Newen & Vogeley, 2003). Essentially, he argues against the widespread view that consciousness has to include self-consciousness (self-reflexive states), and that it has to be connected with propositional attitudes and intentions to act. We want to comment that mental disorders should not only be analyzed with respect to raw or affective feelings, instead we want to stress that in addition to these raw feelings, high-level cognitive emotions also considerably contribute to mental disorders (for the difference between "basic emotions" and "cognitive emotions" see Zinck & Newen, 2008). Often emotional experiences do not only involve basic phenomenal experiences, but are also essentially shaped by

cognitive attitudes such as beliefs and expectations about social relations (e.g. partnership).

Related to the cross-species approach is of course the question as to how the interspecies relation can be bridged concerning mental phenomena. With respect to this "problem of other minds," the traditional view is that we are able to "read" other persons' mental states because of a principal similarity of another person's mental experiences and my own. The most important issue in this respect appears to be the finding that children already understand other subjects long before they acquire a full-blown "theory of mind." To account for that fact, we follow the proposal of Gallagher (2005) who distinguishes between primary intersubjectivity comprising non-conceptual forms of social understanding and secondary subjectivity. This account is in concordance with Panksepp's claim that focuses on non-conceptual (pre-reflexive) consciousness to systematically investigate the neural processes of primary intersubjectivity.

This is related to a basic question that is addressed by all three contributions, which is the question of whether these processes are subpersonal or personal in nature. In contrast to S&G, Panksepp and we argue that one cannot exclude subpersonal or "unconscious" processes in the explanation of mental disorders. As pointed out by Panksepp, information-processing functions are presumably guided by ancient energetic-emotional-affective brain functions. It is plausible to assume that we can only reach a complete understanding of mental diseases after incorporating affective components and cognitive components into a common framework. In accordance to Panksepp, we prefer to develop an informational-processing account to characterize the complex interaction of conscious and unconscious emotions and cognitions. It is, however, by no means clear whether this should be a Neo-Freudian neuropsychoanalytic account.

A more fundamental issue covered in all three contributions is the relation of mental and neural phenomena and subsequently the relation between psycho- and neuropathology. In concordance with our position, Panksepp follows a non-reductionistic monistic view or a dual-aspect theory proposing that neither the first-person-experiential nor the third-person-scientific perspective provide a complete explanation of mental phenomena. We argue for an account that is explanatorily non-reductive, but ontologically reductive, i.e. psychological and neurobiological explanations are non-reductive, but they refer to one and the same monistic world of physical phenomena. A complete description of the functional role of a mental state does not only include neural processes but also inner experiences of the cognitive system and the relation to the external situation. This relation between mental or physical processes can be either spelled out in terms of an identity or a functionalist theory (Newen & Cuplinskas, 2002; Vogeley, 1995). We ourselves prefer the latter because it can account for multiple realizations of the same class of mental phenomena that might be instantiated in different ways in different individuals. However, for one individual or a group of individuals with similar brain organizations, we may be able to characterize mental phenomena in terms of a type-to-type variant of the identity theory. So our functionalist view is compatible with a domain-restricted identity theory.

Although S&G appear to share with us the general position claiming an epistemological non-reductionism with the core idea that a complete description and explanation of mental disorders is dependent on a description in psychological terms, they underestimate the potential contributions of the neurosciences for understanding and explaining mental diseases. With respect to mental illnesses, S&G focus on the so-called "consciousness thesis" that proposes that mental illness can only be properly understood as a disorder of consciousness and that mental disorders cannot be explained without reference to consciousness and psychological terminology. Subsequently, therapy can only be successful on the grounds of conscious concourse. Consciousness is characterized by four distinct features: first, the occurrence of conscious experiences involves the experience of being conscious-of-something, also referred to as "intentionality," second, conscious experiences can themselves be objects of consciousness, third, consciousness may involve extended episodes of sequences of conscious experiences, and fourth, the talk about consciousness refers to the characterization of dispositions, but relevant here are only those abilities whose exercise involves conscious experience.

Their core intuition is that diseases either have a clear bodily or somatic nature (e.g. polycythemia vera) or a mental nature (e.g. schizophrenia), thus implicitly introducing an ontological dualism: if a disorder is of a mental nature we won't be able to identify an organic course of this disease. Such an

identification of a somatic cause of a mental disorder would in fact be a misclassification, thus leading to an exclusive view on mental and somatic disorders, which would subsequently imply that it is not a mental disorder we propose. Even if we had a medical treatment such that schizophrenia as diagnosed today would lead to a complete recovery, this would still not make the psychological description of the disorder superfluous. The description of the behavior of patients suffering from schizophrenia remains essentially dependent on psychological terms even if we would discover a complete "bodily" cause. This is the critical question that makes the difference between our view and their position. This has strong consequences for the conceptual differentiation of "organic" and "non-organic mental disorder" in operationalized classification systems such as the DSM–IV. Conceptually, we do not agree, as S&G do, with this distinction because mental disorders in our view always have an "organic" substrate and a "non-organic" facet of subjective experience that describes how it is to be in a particular conscious state at a given time. Our view thus corresponds to the revisionist position of classical psychiatry that the authors cite as the "somatic basis hypothesis." However, the terms "mental" and "mental illness" cannot be replaced by "neurological" or "brain disease" because the latter terminology does not adequately reflect the world of subjective experiences. Nevertheless, the two terminologies correspond to each other because they both refer, philosophically speaking, to the same extension; that is, the brain with all its neural processes in a given environment.

Following this we clearly disagree with the statement that a sick mind might be related to a healthy brain. In contrast, we propose that mental dysfunctions must correspond to some sort of brain dysfunction. For instance, there is nothing wrong with a proposition such as "Drug addiction represents a disorder in the brain reward system." The comments S&G make at this point are related to problems of granularity of description of neural phenomena. Of course, these brain systems are "unspecific" in that they subserve a variety of psychological functions: the reward system is involved in the complex phenomenon of drug addiction but is, of course, also involved in other non-pathological conscious experiences such as a desire. However, the phenomenon of a desire is also similarly "unspecific" because it can occur under pathological conditions (drug addiction disturbance), as well as under healthy circumstances. This corresponds to a mereological fallacy: we cannot ascribe a particular psychopathological syndrome that is adequately described only on a personal level to a particular brain system that is isolated from the context of other neural processes of the suffering person. Instead, we have to consider brain systems always as a part of a complex set of systems that constitute a nervous system of a human being. Pathology presumably does not involve only one particular brain system, but might be related to an imbalance of different brain systems. This is presumably a matter of degrees of disturbances and balances between different interacting subsystems of the brain, which are not yet fully understood. However, from the fact that we still do not have neuroscientific data, it cannot be inferred that neuropathology can never explain mental disorders.

It is thus non-adequate to infer that mental disorders do not qualify as neurological disorders. In fact, it is the only plausible position that they correlate with dysfunctions of the nervous system. The authors reject this view as "parasitic" because neurobiology would be forced to refer to normative standards for proper psychological functioning that are to be correlated with neural processes, whereby implying an implicit, strong and implausible dualistic account with a certain hierarchy between the mental and the neural. We rather argue that the mental world from which we can be taught about normative values can be of great heuristic value for neurosciences. Moral decisions and cultural differences must also be realized in the brain. That we do not understand much at the moment about culture-dependent differences in cognitive neuroscience does not argue against a monistic view in principle. All mental phenomena have a neural correlate. If this is so, we have to incorporate cultural contexts, individual dispositions and so forth into a complete picture of the neural instantiation of mental phenomena in a given environment. This is exactly the interesting point, but this proposal of course increases the interest for brain functions instead of eliminating them.

According to S&G mental disorders are defined by the fact that consciousness is essentially involved in the cause, the etiology and also in the treatment of the disease. Of course, we accept that cause, development and treatment are important dimensions for the definition of a nosological concept, but the key factor concerning mental disorders is the degree of deviation

from standard psychological competences: *A disorder of a person is a mental disorder if the person has psychological states, processes or abilities which deviate essentially from the standard psychological states, processes and abilities, such that this results in an important change of the conscious experience or the personality traits*. This definition has the advantage that it takes into account that (1) a standard mental disease (e.g. schizophrenia) still remains to be a mental disorder even if we would be able to discover a complete physical cause and an adequate medical treatment, and (2) that mental disorders are essentially culture-dependent. To define mental disorders along deviations from standards that are determined by the society or a culture by no means contradicts that this has an organic correlate of the disorder.

Using descriptions from two different terminologies that reflect a subjective view on the conscious experiences and an objective view on the cognitive and neural processes underlying consciousness does neither imply a methodological nor an ontological dualism. There is still room for integrating the two different perspectives. We agree with the explanatory non-reductionism, but not with the ontological claim. In our opinion, the challenge here is not to weaken the explanatory value of neuroscience. Rather, it is to strengthen the view that mental disorders can only be further explored with the deep understanding that these diseases are involving different levels of description, which have to be integrated in an approach of multi-level explanations that are not competitive but cooperative in nature. In summary, we defend a monistic account including epistemically non-reductive explanations. But these different explanations could be essentially connected and integrated with each other. Conceptually, it is most important and fruitful to transfer mental illness to brain disorder because this is a major strategy to understand the pathophysiology, including its genetic basis, of diseases and to provide new strategies for therapies.

References

Gallagher, S. (2005). *How the Body Shapes the Mind*. Oxford: Oxford University Press.

Newen, A. & Cuplinskas, R. (2002). Mental causation: a real phenomenon in a physicalistic world without epiphenomenalism or overdetermination. *Grazer Philosophische Studien*, **65**, 141–169.

Newen, A. & Vogeley, K. (2003). Self-representation: the neural signature of self-consciousness. *Consciousness and Cognition*, **12**, 529–543.

Vogeley, K. (1995). *Repräsentation und Identität*. Duncker und Humblot.

Zinck, A. & Newen, A. (2008). Classifying emotions: a developmental account. *Synthese*, **161**, 1–25 .

Response to commentaries

G. Lynn Stephens and George Graham

Are we dualists? Kai Vogeley and Albert Newen believe that we are. This is because we talk in our target chapter of the roles of conscious experience in mental illness as if (to them) we intend to refer to something non-neurobiological. Also, they ask: when we say that a disordered mind may be embodied in a healthy or normally functioning brain, aren't we denying that mental illness is physical? And, finally: Are we more or less neglectful – as Jaak Panksepp worries – of the roles of emotion, mood and affect in mental disorder? Do we wish to downplay those roles?

These are three significant, if not the only questions posed to us by our section colleagues in their commentaries. Below we address them in order. (1) Suppose that each and every mental state is somehow nothing but a type of brain state. Call this a version of The Base Thesis. This is the thesis (mentioned in our target chapter) that mind and mental illness has a neurophysical base. There are other versions of the base thesis, but this particular version presupposes that the metaphysical thesis of physical monism is true. Other versions do not. What follows from this monist inspired version of the base thesis – call it the Basal Identity Thesis (BIT) – about how mental illness is best understood and explained?

Not much, we claim. Metaphysical monism does not entail explanatory monism.

We claim (in CT, our consciousness thesis) that mental illness is illness in and of conscious experience or activity, broadly understood. By this we mean that it has to be understood and explained in robust part in terms of the language of conscious experience. The language of neuroscience by itself is insufficient to understand and account for mental illness even if BIT is true.

Consider the following analogy. A physicist from the Martian Institute of Technology (MIT) lands on planet earth and observes a baseball game at Yankee Stadium. He knows a lot about physics, but nothing about the rules of baseball. Each and every move of each and every player in the game is physical. Can physics fully explain why the players act as they do? No, of course not, for while each player is in motion (and physics addresses that fact), not everybody *merely* is in motion. The players are following the rules of the game and their actions cannot be understood without reference to those rules and to each player's knowledge of them.

Now suppose that someone suffers from a mental disorder, say, alcohol addiction. (Assume, for the sake of illustration, that alcohol addiction is a mental disorder in the manner identified by CT.) Can the Martian understand why this person consumes alcohol as imprudently as they do? Not, on our view, unless the Martian also identifies various relevant conscious events and experiences of the victim as well as the explanatory roles that such events play in the disorder. Our saying such things about mental illness, our privileging consciousness, does not make us metaphysical dualists or anti-monists. It makes us realists about mental illness. If mental illness is real honest-to-goodness illness and is distinct from brute somatic disorder ('brute', meaning that consciousness plays no critical role), it has to be understood, in part, in phenomenological terms. A neurophysicist from MIT observing an alcoholic in action needs to know the 'rules' or conscious experiences and tendencies responsible, at least in part, for the disorder. (2) Suppose, again, that BIT is true. Does this mean that each and every mental illness is a brain disease or neurological disorder? No, it does not. We believe that some mental disorders are not disorders of the brain even if BIT is true. Consider the following analogy.

An inexperienced baker is baking bread in an oven. He wonders what the temperature of the loaf is. So, he foolishly takes an oral thermometer from his medicine cabinet and sticks it into the piping hot loaf. The thermometer's mercury rises immediately and bursts the device. Does this mean that the thermometer is functioning improperly? No, the answer is, not at all. Not every burst of an oral thermometer is of a thermometer that fails to function as it should. The baker used his thermometer improperly. It's meant to be placed under the tongue and not in a hot loaf.

Alcoholics behave improperly. Mother Nature designed the brain to regulate consumption, learning, memory and much else besides. The brain may also help to produce addiction, just as a thermometer may be deployed to determine the temperature of a hot loaf. However, when a person is addicted, this does not mean that the brain is failing to behave as it should, any more than an oral thermometer is at fault in a hot loaf. The person is not behaving as *he* should. He is suffering from a mental disorder that is not also a brain disorder, even if the disorder is a physical state of the brain. It can be a physical state without also being a disordered physical state. Again: Not every mental disorder is a disorder of the brain, even if every mental disorder is physical. (3) We regret that we do not have space here to discuss our views about the role of emotion in mental disorder. In our commentary to Jaak Panksepp's target chapter, we did not wish to convey the impression that we regard that role as relatively less important or weaker than that of the cognitive aspects of mental illness. For one, we believe that the conscious experience of stress may tilt (to use a metaphor of J. Allan Hobson) the brain's chemistry towards "emotional problems acting as triggers for a wide range of mental problems" (Hobson & Leonard, 2001, p. 239). Elsewhere we have written (though not in detail) about the role of emotional stress and other aspects of emotion in mental disorder (see Graham & Stephens, 2007). Interested readers may wish to consult that paper. That paper also contains complementary discussion of the points we make in (1) and (2) above.

One final point with which to close: nothing is "monolingual" about understanding mental disorder. From a medical-scientific point of view, it is eminently desirable that all sorts of information from neuroscience contribute to our understanding of mental illness, even though that information, given the truth of CT, is ultimately incomplete. So, for example, if certain stress-related memories (to use the language of consciousness), stored in the amygdala (to use a term from neuroanatomy), but normally inactive, can in certain settings become active and set off fight or flee responses (to use the language of behavioral neuroscience that applies as well to nonhuman animals), then such a complex fact may help to explain the onset of debilitating anxiety or phobia. At their individual bests, such multiple languages or schemes cohere like close brothers and sisters. Each helps with certain explanatory responsibilities that their siblings cannot discharge on their own.

References

Graham, G. & Stephens, G. L. (2007). Psychopathology: minding mental illness. In P. Thagard (Ed.), *Philosophy of Psychology and Cognitive Science* (pp. 339–367). Amsterdam: Elsevier.

Hobson, J. A. & Leonard, J. (2001). *Out of Its Mind: Psychiatry in Crisis: A Call for Reform*. Cambridge, MA: Perseus.

Understanding affects: toward a neurobiology of primary process mentalities

Jaak Panksepp

One of the most infamous statements in biological psychiatry is that "Behind every crooked thought there lies… a crooked molecule" (see Pincus (2000) and Slater (1999) for the context of this famous quote by Ray Fuller, one of Prozac's co-inventors at Eli Lilly). This kind of thinking neglects the importance of lived lives in managing psychiatric distress. This resembles the concern raised by Stephens & Graham about my advocacy for a neuroscientific approach to affective consciousness, which reflects both my main scientific and psychiatric concerns. They wisely question whether anything about the anatomy (and presumably other detailed features) of brain emotional systems "will help us to decide between competing accounts of the origins of, say, anxiety disorders?" My answer to this is "yes" and "no." Brain facts will certainly not help us fathom the individual life trajectories that have led an individual down the path of disabling anxiety. However, they may help us generate a better *evidence-based* psychiatry (with all the equivocations we must have about such laudable goals – see Little, 2003).

For instance Don Klein (1964) provided an early example of how two anxiety syndromes, panic attacks and generalized anxiety disorders, could be distinguished on the basis of one responding much better to tricyclic antidepressants while the other was quelled selectively by the early benzodiazepines. Although this rule of thumb has been muddied by the utility of more potent benzodiazepines (i.e. alprazolam is effective in panic), it highlights the multiple neurobiological paths to anxiety, an active field of animal investigations (Vianna & Brandão, 2003). There are probably life history and phenomenological differences among anxieties that correspond to brain differences. For instance, different anxieties emerge from FEAR and separation-distress/PANIC systems of the brain (Panksepp *et al.*, 2009). Also, as these systems *sensitize* from traumatic vicissitudes of life, the changing morphologies, gene-expression patterns, and neurochemical titres – the substrates of primary-process mentality – should be of considerable importance for psychiatric understanding and effective management of emotional distress. The biological views and life trajectory views surely need to work better together.

I thank Newen & Vogley for their insightful synopsis of my views on cross-species "subpersonal" aspects of emotional life. I affirm their call for scientific approaches to psychiatrically relevant consciousness studies that face up to the enormous empirical challenges posed by primary-process brain mechanisms. I find their preference for a "functionalist" over my so-called "identity" account of affective processing to be an appropriate move for the higher cortico-cognitive aspects of emotional processing, but perhaps not as useful for understanding evolutionarily provided emotional tools found at the subpersonal, subcortical level. Those modes of processing may blend within convergent-integrative higher limbic zones.

Can we agree that a most critical scientific issue for all of psychiatry is clarification of the biophysical nature of primary-process affective experience? There are hardly any neuro-mechanistic proposals in the area (but see Panksepp, 1998). Existing evidence suggests that the most psychiatrically useful biological knowledge may emerge from understanding the evolved subpersonal, subcortical realms (the depth neuro-psycho-biologies of our animalian "souls"), rather than by describing the diverse life experiences of individuals. It certainly appears that our minds are grounded on a very complex and presently unfathomed subcortical core-SELF structure, homologous in all mammals, laid out in viscero-somatic action coordinates. Here is where a great deal of psychiatrically significant distress is probably felt, and where the general principles must be sought for a scientific understanding of affective experience. It may well be that all forms of consciousness are still grounded to the basic values of primary affective

states. If so, cognitions, defined as information-processing states, may derive their psychiatrically significant power from affects and less the other way around.

The evidence-based "identity view," derived from abundant evolutionary-functional homologies within subcortical limbic circuits, encourage us to consider that animal models are the most robust scientific ways to understand the foundations of human affective consciousness. When we accept that we are "just" mammals, the epistemology in this murky area becomes straightforward: as one studies the neural substrates for genetically ingrained instinctual responses, one is harvesting critical knowledge about those subpersonal emotional feelings, deeply experienced by those in psychiatrically significant distress. Within the massive random-access associative spaces of neocortex, a functionalist view has many more degrees of freedom for individualized mental navigation. To the extent that psychiatric disorders reside more in those neuromental spaces, we have less hope (certainly fewer strategies) for deriving therapeutically useful general principles from any fine-scaled neuroscientific analysis, and hence clinical wisdom will prevail.

We must elucidate details of evolutionarily provided subpersonal emotional systems for major progress in psychiatric medicinal development and for understanding the shared substrates of emotional imbalances in psychiatric disorders. This does not contradict the critical importance of individual lives in psychiatry; N&V correctly encourage us to also focus on "the importance of other high-level cognitive emotions for mental disorders" – I did not only because of space limitations and the inability of animal models to access cognitive issues as effectively as basic affective ones. Still, perhaps the best basic science investments in this area will be in decoding the cross-species subcortical affective substrates in animal models, which may offer sufficiently detailed general principles of neural action that can advance biological psychiatric therapeutics. Comparable investments on cognitive issues are less likely to provide robust general principles, if for no other reason than massive cross-species and individual differences at that level of analysis. Might "functionalist" variability in higher brain mechanisms generating comparable psychological effects be a barrier to robust scientific progress?

Many foundational neuropeptides and other neurochemistries remain to be functionally characterized and harnessed for development of more precise psychiatric therapeutics (Panksepp & Harro, 2004). A neuroscientific confrontation with those mysteries should percolate naturally into practical clinical concerns: when we understand the neurochemistries of RAGE circuitry, we may have anti-anger medicines, quite useful for regulating that endophenotype in many psychiatric syndromes. Consider that separation-anxiety is quelled better by opiates than benzodiazepines: since social loss is a major vector in depression, might we wish to consider non-addictive mixed opiate receptor agonists/antagonists such as buprenorphine in the treatment of depression (Watt & Panksepp, 2009)? Preliminary evidence affirms the remarkable efficacy of such approaches (Bodkin *et al.*, 1995).

The issue of *emergence* will continue to haunt us as long as our knowledge remains incomplete. However, if emotion and affect in brain–mind evolution served as the primary-process for all subsequent developments in consciousness, solid scientific progress at higher levels may remain linked to our understanding of brain–mind substrates. I believe that evolutionarily, consciousness likely emerged from the coding of biological survival values (unless we believe in a *panexperientialist* quantum "mind dust" permeating the physical universe). A compelling but vastly ignored idea is that our remarkable cognitive abilities are grounded in core survival themes (primary affects), allowing cognitions to rapidly weigh alternative courses of action that may facilitate or hinder survival. That, I believe, is the ultimate function of affective consciousness, and why affects, as Freud surmised, are never unconscious, while many cognitive activities work effectively when out of mind. Despite the success of cognitive-behavioral therapies, emotional homeostasis will be most rapidly adjusted through a better understanding and use of affective issues, both biological and psychological. Concurrently, clients' cognitive concerns should never be minimized.

How to facilitate emotional awareness, in both clinical as well as scientific practice, remains a momentous challenge for both biological psychiatry and cognitive neuroscience. We have barely initiated systematic inquiries into the "heart of darkness" that affective experience poses for our scientific understanding of lived lives.

I appreciate this most interesting opportunity to discuss topics of momentous importance for

psychiatric thought. I hope the scientific community will pursue, more courageously, illumination of the neurobiological underpinning of affective consciousness. It is finally a do-able task, because emotional feelings are closely linked to instinctual action networks of the brain. By contrast, obtaining *causal* understanding of individual human cognitive experiences remains, regrettably, next to impossible. Perhaps that makes the topic less workable for development of *new* evidence-based medical practices. Psychoanalysis, especially when it comes to be based on modern neuroscience, will remain the most comprehensive way to understand individual mental landscapes.

References

Bodkin, J. L., Zornberg, G. L., Lucas, S. E. & Cole, J. O. (1995). Buprenorphine treatment of refractory depression. *Journal of Clinical Psychopharmacology*, **16**, 49–57.

Klein, D. (1964). Delineation of two drug-responsive anxiety syndromes. *Psychopharmacology*, **5**, 397–408.

Little, M. (2003). 'Better than numbers…' A gentle critique of evidence-based medicine. *Australia and New Zealand Journal of Surgery*. **73**, 177–182.

Panksepp, J. (1998). The periconscious substrates of consciousness: affective states and the evolutionary origins of the SELF. *Journal of Consciousness Studies*, **5**, 566–582.

Panksepp, J. & Harro, J. (2004). The future of neuropeptides in biological psychiatry and emotional psychopharmacology: goals and strategies. In J. Panksepp (Ed.), *Textbook of Biological Psychiatry* (pp. 627–660). New York, NY: Wiley.

Panksepp, J., Fuchs, T. & Iacobucci, P. (2009). The basic neuroscience of emotional experiences: the case of FEAR and implications for clinical anxiety in animals and humans. *Applied Animal Behaviour Science*, in press.

Pincus, D. (2000). Mind and brain sciences in the 21st century. *Psychoanalytic Psychology*, **17**, 600–607.

Slater, L. (1999). *Prozac Diary*. London: Hamish Hamilton.

Vianna, D. M. L. & Brandão, M. L. (2003). Anatomical connections of the periaqueductal gray: specific neural substrates of different kinds of fear. *Brazilian Journal of Medical and Biological Research*, **36**, 557–566.

Watt, D. F. & Panksepp, J. (2009). Depression: an evolutionarily conserved mechanism to terminate separation-distress? A review of aminergic, peptidergic, and neural network perspectives. *Neuropsychoanalysis*, **11**, in press.

Replies to comments by Jaak Panksepp and by G. Lynn Stephens & George Graham

Kai Vogeley and Albert Newen

Comments

Again, we are grateful for the considerations of our colleagues on our contributions that consider both basic aspects of consciousness as well as clinical issues. Summarizing our view on consciousness, a solid concept of consciousness has to cover both conceptual and non-conceptual properties and cannot be confined to a reflexive, "cognitive" domain, but should also include intuitive, pre-reflexive experiences. Panksepp refers to the latter with the term of "raw primary-process experience." What the evolutionary relevance of this "layering" of consciousness might be is an interesting question and cannot be fully answered on empirical grounds yet. In our view, a theory of cognition has definitely high relevance for the understanding of consciousness. In addition, we generally agree with Panksepp that affective experiences need to be incorporated into a full account and understanding of consciousness and its role for mental disorders. Such a pre-reflexive self is not only constituted by affective states, but there are at least the additionally important dimensions of (1) motor activities while acting in an environment, (2) perception-based representations of consciously experienced objects and (3) social interactions with other people.

Let us consider the particular aspect of social interaction in greater detail. In addition to affective feelings, a core factor of the development of mental disorders is social interaction. Panksepp accepts that general claim, but holds that social interaction is essentially determined by fundamental affective feelings. Of course, the interaction with a person is usually presupposing a positive affective feeling towards a person. Nevertheless, the social interaction is only essentially constrained by the affective feeling if those are either extremely positive or aversive. If there are only "standard" affective feelings involved – as it is presumably the case in most of our interactions with others – then the actual interaction is more important and may be the main factor leading to an affective evaluation of the person. It may happen that the low-level negative affective evaluation at the first

glance is changed by the everyday interaction into a strong positive affective stance. The actual interaction includes not only language-based, but also non-verbal communication including gesture, gaze, posture and other essentially embodied means of communications.

We agree with Panksepp that the influence of emotions has to be taken into account to understand mental disorders, but there is of course again the danger of reducing mental disorders to affective disorders. In his comments Panksepp is making a quite strong claim, namely, that affective forces drive cognitive processes. We are suspicious whether emotional experiences have always been the root for cognitive processes. On the one hand, there exist alternative theoretical accounts that try to reconstruct emotions as cognitive processes (as proposed for instance by Antonio Damasio). In our own work, we also argue that we have to distinguish basic emotions that are essentially independent from complex emotions that involve high-level cognitive processes. The latter are classified as primary and secondary cognitive emotions (Zinck & Newen, 2008). On the other hand, there are concepts under debate that try to reconstruct the core deficit of psychiatric disturbances on the basis of fundamental information processing, for instance organization of behavior in the temporal domain (Vogeley & Kupke, 2007). We are in favor of a multifactorial concept of mental disorders that investigates the interrelations between the important factors of emotion, perception, action and abstract thinking. We are of the opinion that neither the cognitive abilities nor the social interaction can be reduced to fundamental affective states. To fully account for mental diseases we have to investigate the complex interaction between affective, cognitive and social competences as they are realized by neural processes. To reach this aim we should develop a close connection between neuropsychiatry of adults and the developmental psychology and cognitive neuroscience investigating the ontogenesis of children and the genesis of mental diseases.

With respect to the clinical domain, Stephens and Graham (S&G) point out that "a patient's problems with 'self-other differentiation' are quite different" within different diagnostic groups. They further propose that these "differences indicate the scope and complexity of problems that must be faced in trying to understand how people distinguish their own mental states from those of others or may fail to do so in cases of mental disorder."

We are grateful for the fine-grained considerations that are pointed out by S&G with respect to a necessary distinction with respect to self-other differentiation and self-other exchange. Let us first consider disturbances of self-other exchange that cover the phenomena of delusion and social cognitive disturbance of autism according to our concept. Chris Frith has recently put forward a very fruitful distinction separating "hypermentalizing" and "hypomentalizing" (Frith, 2004). The best example for the hypermentalizing disturbance is the phenomenon of delusion as a first-rank Schneiderian symptom for schizophrenia, during which more than necessary information is read out from random datasets which are usually not considered informative during non-disturbed ("ortho"-)mentalizing processes. A specific phenomenon might gain extraordinary importance during the delusional experience such as a black shirt of another person that might indicate sudden death. According to our view, delusional experiences are disturbances of the ability of self-other-exchange or perspective taking. The person suffering from delusional experience is no longer able to share his or her experiences with others; this experience has become highly private and rigid and can no longer be corrected by others.

In contrast, hypomentalizing is true for the diagnostic group of autism in which mentalizing is weak or impossible. The disturbance or inability to take the perspective of others ("standing in someone else's shoes") is the most prominent symptom of autism that occurs along the entire spectrum of the disorder (low- and high-functioning autism). Autistic persons have difficulties in imagining what other persons think or feel, in understanding implicit messages (including irony, metaphors etc.), and non-verbal communication. This disturbance or inability has severe consequences for their social life.

The phenomenon of thought insertion is clearly different. Thought insertion as another Schneiderian first-rank symptom besides delusions belongs to the group of phenomena covered by the term ego-psychopathology ("Ich-Stoerungen"). In our view, this is a typical example for a disturbance of the ability of self-other differentiation that usually allows us to ascribe a mental phenomenon to oneself or to other persons. This ascription is obviously no longer adequate if a mental phenomenon such as a thought is ascribed to someone else. To fully account for this phenomenon we need to distinguish affective and cognitive levels in a two-step account that distinguishes

a disturbance of the feeling of ownership from the cognitive judgment of ownership (Vosgerau & Newen, 2007). Meanwhile the analogous distinction of feeling of agency and judgment of agency is relevant to account for delusions of control (Synofzik *et al.*, 2008).

The interplay of self-other differentiation and exchange is definitely an intricate complex of different cognitive and affective processes that need to be balanced in a subtle and fine-tuned way, in order to allow everyday interaction and communication with others. Functional imaging studies focusing on self-referential and social cognitive processes suggest that both processes recruit similar brain regions including the anterior medial prefrontal cortex and the temporoparietal junction. The fact that these regions are also part of a so-called "default mode of brain function" (Raichle *et al.*, 2001) that correlates with resting activity might be an indication of the fact that our brain has a natural disposition for these self-referential and social cognitive processes. This, however, is yet only a speculative hypothesis and sets up another research agenda that must be followed up in more detail in future research programs.

References

Frith, C. D. (2004). Schizophrenia and theory of mind. *Psychological Medicine*, **34**, 385–389.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z. *et al.* (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences USA*, **98**, 676–682.

Synofzik, M., Vosgerau, G. & Newen, A. (2008). Beyond the comparator model: a multifactorial two-step account of agency. *Consciousness and Cognition*, **17**, 411–424.

Vogeley, K. & Kupke, C. (2007). Disturbances of time consciousness from a phenomenological and a neuroscientific perspective time. *Schizophrenia Bulletin*, **33**, 157–165.

Vosgerau, G. & Newen, A. (2007). Thoughts, motor actions and the self. *Mind and Language*, **22** (1), 22–43.

Zinck, A. & Newen, A. (2008). Classifying emotions. *Synthese*, **161**, 1–25.