

Draft

Norms and Neuroscience: the case of Borderline Personality Disorder

In a recent retelling of an Aboriginal story, a creature emerges from a swamp without any idea of what it is (Wagner & Brooks, 1977). A platypus tells it that it is a bunyip, but learning the word does not go far to answer the creature's questions about itself. The bunyip then discovers a scientist, a very busy one, who without looking tells it that buniyps do not exist. Depression ensues until the bunyip discovers another one just like it.

The bunyip's encounter with the scientist might seem to capture a justified fear that many feminists have felt about the scientific faltering gaze, a gaze still typically male. Far from an insightful examination of how women function, the male scientist may take himself as the model of the healthy human being, and then find healthy women do not exist (Potter, 2009; Wirth-Cauchon, 2001). Such a picture of scientific psychology and psychiatry is of course an exaggeration, nor is the feminist view accurately reported here. But there is enough in it to lead many to expect a discussion of 'Borderline Personality Disorder' from a neuroscientific perspective is hardly likely to get much right.

It would be unfortunate to miss out on recent neuroscientific work on BPD. There is a new understanding of BPD that derives from such research (King-Casas et al., 2008; 2005). The result is quite revisionary; it challenges previous clinical and ordinary understandings of what the behavior of those with BPD signifies. Most of our attention will be concerned with the new interpretation, but we will start with an explication of what BPD is, and an objection to the classification. The neuroscientific research weakens this objection and so it and its relation to philosophical understandings of CSN are the focus of this paper.

I am hopeful that the following essay will at least introduce an alternative and acceptable way to view some of the field's findings. To do so, we will need to sort through conflicting viewpoints. One motivation for doing this work is that our setting is very different from that of the bunyip. Far from passing through our environment scarcely described, most of us are told almost continuously what we are, what we are doing and why we are doing it. And, of course, what we should be doing.

Borderline Personality Disorder

BPD can involve relatively distinctive behaviors, some of which look very much like tantrums directed to someone who is hardly deserving of such treatment. An example

involving a psychiatrist who accompanied her patient to a meeting about the results of removing cancerous tissue (Carol W Berman, 2014):

That day I went into the ob-gyn's office with her and sat across from the doctor who reported great news that the patient was cancer-free. ... Out in the hallway... my patient yelled and cried.

"You colluded with her! I can't believe how you doctors were so self-satisfied. You didn't even consider me. You and that doctor talked down to me like I was a moron!"

... "I hate you both!" she screamed and ran down the hall. I dashed after her, calling her name, but she jumped into an elevator and ran off.

Theorists, social critics, and psychotherapists who write about BPD tend to agree that it is typified by a number of features, including some or all of the following: an unstable sense of self; a deficit in empathy; a fear of abandonment and an inability to sustain relationships; risk taking and self-harming; splitting or successively idealizing and damning individuals. There are sometimes differences in emphasis, with the fragility in self accorded predominance in some accounts and the destructiveness of relationships emphasized in others. It is in fact possible to map out neurological deficits in many of these categories (Friedel, 2004) .I am going to take the one discussed below as having a strong case for being fundamental. The case consists in the fact that it shows us the BPD subject has very serious deficits in social cognition, where having accurate understanding is very important in the acquisition of many social skills, including emotional regulation, a sense of social self, and so on.

From a political point of view, the typifying traits we have just seen can also easily appear to be derived from a set of behaviors that many women, at least in the West, arrive at as the best of a very narrow set of options allowed to them through the restrictions of their society. In an environment where a women's capacity to rationally select among options is denigrated or denied, an emotional outburst may remain as the most effective instrument to influence others. When a woman's worth, including her very livelihood, depends on her relationships with others, abandonment may indeed be greatly feared.

Those who object to the BPD label on feminist social constructionist grounds may be starting from just this point of view. According to a social constructionist account of some phenomenon, it is created by society and culture, as opposed to occurring naturally, independently of the specifics of a culture's beliefs and evaluations. A feminist version of the view tends both to be concerned with the effects of being labeled in term of the phenomenon and to emphasize the power relations involved in using the labels.

Accordingly, a feminist may well approach the use of 'BPD' as a diagnosis with the expectation that it reflects not some well-grounded individual psychopathology but rather the negative results of following the choices more or less forced on women. And added to this is the failure of women to fit a male model of independence and cohesion. We can also include society's great discomfort with uncooperative women, such as those labelled 'witches'. We might, then, think that the unattractiveness of BPD is really in the eye of

the beholder; as such, it is evidence of a social creation. However, recent research tells us that those with BPD are deeply incapable of establishing the cooperative relationships so many of us can enter into. They are not capable of having the perceptions of norm violation that are critical to cooperative behavior. This result takes BPS out of the domain of social invention.

A Neuroscientific Approach

The woman in the episode described near the opening of the previous section seems highly sensitive to slights. At the same time, we might well think that she cannot fully comprehend how hurtful her reaction to the therapist would have been. But in fact the science reverses this story (King-Casas et al., 2008; 2005)

Subjects were asked to participate in a well-studied “trust game” (Berg, 1995). A 'trustee' and an 'investor' successively give each other some of money. Success depends on trust and cooperation, including being able to repair the trust if broken. The BPD subjects in particular could not repair broken cooperation, in part because they could not read the signals that the trustee sent to indicate mistakes on their part.

The explanation of the failure comes from fMRI scans that reveal an asymmetrical reaction in one brain part, the anterior insula. The insula is involved in reactions to norm violations, both when the subject violates a norm and when the subject is the target of a norm violation. According to the readings of the fMRI scans, the BPD subjects were registering when they were violating norms, but they did not experience the norm-violating negativity directed at them. This result is contrary to what clinicians and others ordinarily think is going on with BPD subjects. That is the idea that BPD subjects are vulnerable people with challenged egos who fly off the handle when they perceive a slight. It is much more accurate to say that BPD subjects do not perceive slights.

Let us agree that the neural site of the signaling deficit is the anterior insula, which failed to discriminate ordinary from norm-violating input. Can we take this as the basis of BPD, its ground or even part of its essence? If this works, we can bring in a rich account of concepts to articulate a structure for our understanding of borderline personality disorder. The account, sometimes called the theory account (Machery, 2009; Murphy, 2002), takes concepts of kinds to contain both recognition conditions and also a reference to a underlying cause of such conditions.

We often enough take the underlying cause of recognition conditions to be what the kind really is. This is obviously so with simple substances, such as gold or water. Water, to simplify matters quite a bit, is H₂O. Can we extend such an account to psychological kinds? Could we take the disposition to have asymmetrical reactions in the insula to be what borderline personality disorder really is?

The analogy with water has some further facets. The recognition characteristics associated with 'water' and its chemical composition fall short of what counts as anything like an adequate understanding of what water is. For example, consuming water is vital to human beings, and many other forms of life. Without such knowledge it would be hard to understand our interest in water. However, such uses by life forms might well not count as essential to water. A loss of life on earth would not as such knock out the presence of water.

It has been argued recently that psychological illnesses, in contrast to simple chemical substances, are essentially connected to aspects of the society in which their sufferers occur (REF). On such an account, BPD is not and cannot be purely neural. Without any manifestations, there is no BPD. Without disputing this claim, we can instead focus on the dispositional aspects of BPD. Dispositional-BPD may or may not be identical to full-BPD. Whether it is should not matter to the question we are heading towards: is either identical to some neural condition.

At the same time, we should stress that human interests in BPD, like human interests in water, go vastly beyond anything like neural conditions. BPD is in many ways a social phenomenon, rather as water is an economic phenomenon. But our attention will be on the neural basis of BPD. A major reason for this attention is just the already noted fact that recent work requires us to revise our understanding of BPD. The exceptionally important manifestations of the condition may mislead us about what is, in some sense, really going on.

Philosophical Objections

There are a number of sources of objections to any type of identification of BPD, or part of it, with brain activity, either actual asymmetrical activity or a disposition to such activity. We will concentrate on two; the first we will look at is the insistence on a personal/sub-personal distinction (Dennett, 1969; Drayson, 2014; Hornsby, 2000). The second comes from the reductionist arguments we have from Kim (Kim, 2005, 2007; McLaughlin & Cohen, 2007). There are other grounds for contests of identity. The strongest is the argument from multiple realizability, which has received important attention; nonetheless, our view will be that it has also recently been too problematized to be employed in any short way here (Polger, 2016).

The personal/sub-personal distinction is sometimes presented as just the observation that properties of a whole person cannot be attributed to a proper part of the person. E.g., Jones believes, but Jones brain does not believe, even if Jones brain's activity is thought to be the main site of belief-related neural activity. There are, however, too many cases where part-whole boundaries are not respected in the attribution of properties. Thus, if one says one is in pain, the question "Where does it hurt?" is an invitation to locate the pain in a part of the person even though being in pain is attributed to the person. In addition, for a number of theorists, psychological states are to be understood in terms of

brain mechanisms, where descriptions of the contribution of the parts are very important to our understanding of the grounding of the states. For such theories, descriptions of the mechanisms can provide a complete characterization of the psychological state.

A more seemingly solid version of the objection starts by saying that personal level properties are often normative (Thornton, 2013). But the level of neural facts investigated by cognitive neuroscience is not normative, it is maintained, and so it cannot constitute the essence of personal facts. This objection can be connected with the Kim objection: Given a separation between normative and neural properties, the Kim objection can be put very swiftly. Physical properties must have sufficient physical causes and normative properties cannot be additional causes of physical properties because of the causal exclusion principle, which states that one cannot have two individually sufficient causes for some physical condition. Similarly, normative, personal level properties cannot be (partially) constitutive of most neural reactions, and vice versa.

We can see proponents of extended embodied cognition as developing a third version of the argument from the personal/subpersonal distinction. For them the problem is that subpersonal causal relationships are not going to add up to some personal level phenomenon. I think we can agree in part with this point. For example, one needs far more than asymmetrical activity in the insula to get the chaotic relationships borderline persons can have. But just as the biological importance of water does not militate against a chemical definition of it, so it seems right to say that the consequences of the asymmetrical activity do not show that BPD cannot be identified with a neural condition.

To use terminology some will be familiar with: We might think of BPD as a whole constituted by neural conditions and their consequences. If we do this, then it seems right to say that the consequences actually partially constitute the whole and are not literally caused by it, though they are caused by a part of it. On the other hand, we may be able to think of BPD as actually constituted only by neural conditions, with the rest positioned as effects. Whether there are large obstacles to the latter is what we will continue to look at.

A Defense

The question of levels and reduction is often envisaged as involving levels where, as in Hume's world, one little thing happens after another. The little things may be at the atomic or neural level, but on this picture, the explanatory aim is to uncover regularities and the laws that express them. A normatively neutral description in terms of mechanisms is presumably also possible. Instead of one little thing happening after another, we have little things interacting with other little things, or parts of them. Instead of laws, we have algorithms for interactions. Whether or not such a picture fits physics, it drastically mischaracterizes CNS. Cognitive neuroscience has an explanatory task that is quite different. To see this, we can look at a norm-neutral model of a cognitive neuroscientific explanation:

But one sort of understanding that cognitive scientists are often interested in achieving is analogous to the understanding that one would have of a clock if one could identify each of its functional parts (its springs and cogwheels, its pendulum, and so on), and the way in which all these parts interact to bring it about that the clock has a reliable disposition to tell the correct time...

...An analogous understanding of how a computer works would involve an understanding of the structure of its electrical circuits and of the logical structure of its programming code. If this is the sort of understanding that cognitive science is particularly interested in, that would help to explain why cognitive scientists are so interested in actually trying to build machines that can do some of the things that minds can do.

Thus, at least one of the goals of cognitive science will be to explain the micro-level processes that are characteristic of the mind. These are processes in which one mental event or state is caused by another mental state or event that precedes it as closely as one mental event can precede another...(Wedgwood, 2006)

It is important to see, nonetheless, that this is not true of cognitive science as actually practiced, and that it should not be true of it. We can see this by noting that no amount of internal investigation can by itself tell us how a clock has a reliable disposition to be correct. The idea that examining the internal workings of a clock will help us understand how it has the disposition to tell time has two large problems on its surface. First of all, a clock ticking over may not have that disposition. Perhaps it runs fast or slow, or perhaps its running is irregular or intermittent or both. Secondly, if you decouple the watch from its environment and take it away, it can stop telling the correct time without any internal change.

The clock example indicates that philosophy's aims are different from those of CNS. Philosophy is often interested in where the line between, e.g., seeing and not seeing is drawn. CNS, in contrast, is interested in what it is to see well. The central task of CNS, as Montague and Quartz have pointed out, is to explain how a creature flourishes in its niche (Montague & Quartz, 1999).

Thus a neuroscientist looking at the organs we employ in cognitive activities is looking at those organs' function and how they perform that function. Performing their function will be part of what it is for the organism to flourish in its niche. But given that this is what introduces the normativity, cannot we just naturalize "function" and so eliminate the normativity?

Many philosophers have advanced the idea of naturalizing "function" and "functioning", but it is not clear how successful they have been. To decide the question of their success we have to decide what they need to do in the present context. And that is to provide an account of "well functioning" that shows it is not an irreducibly normative concept.

Our discussion below will mention views about the role of evolution, and so it is well to remind ourselves of some ways in which evolution does not loom large in cognitive neuroscience. For one thing, there are features of human functioning where we just do not know whether they are learned or inherited. Our vision has features particularly suited to creatures that live in a world where a major source of light is from above them. For example, what well-sighted human beings see as convex in contrast to concave depends on where the shading is, which is a consequence of the direction of light. But even though the influence of the shading reflects the standard light source, we do not know whether these features are acquired or inherited (private communication with Dr. Harold Bedell). Somewhat similarly, almost all of us have vision that can function well driving at night. In particular, lights from passing cars are not experienced as extended blurs, which could be very misleading and dangerous. Single moving lights can easily produce a blurred experience of, for example, a circle of light. Our normal night vision is quite likely produced by backward masking. It is very unlikely that evolution was sensitive at all to the demands of driving at night, and we do not learn or otherwise acquire the use of masking at night. Rather, we build environments that enable us to function well at night (private communication with Dr. Haluk Ogmen).

There have been two particularly significant accounts of “function” developed in the literature. One, initially proposed by Robert Cummins, draws on the idea that the function of some trait is a matter of its causal contribution to the larger system of which it is a part. There is, however, an obvious problem with this definition; a trait or thing will end up with a very large array of functions, since it will be part of multiple systems. In the current response, the effects that it is a thing’s function to produce depends on which systems we are interested in. That is, it functions to produce the effects for the system that we are interested in. There are a number of different problems with this account, but for our purposes relativizing a function to our interests is not a way to naturalize the normativity in “functioning well.” “Our interests” bring in both perspective and normativity. They are a matter of what we think is important.

A second and seemingly much more promising account is the selective effect account. On this view, functions of objects or traits are selected by evolution. Functioning well is just then functioning as trait or thing is selected to function. And “selected” here may just mean that it is that functioning that contributed to the trait or organism increasing its presence in a relevant population. For this view, we can think of the items as having numerical success, which is to be understood in factual terms as increasing its representation.

There is a problem with this view, however. The problem is that there are ways of functioning that come on the scene too recently for evolutionary success to have had much role. The three examples I will use come from reading and writing, driving automobiles at night, and ballet. Each of these is late in evolutionary time, with some relatively quite recent. Each of these also requires and employs neural resources that evolved independently of the benefits of the later functions. Our ability to read and write seems not, therefore, to be the result of some evolutionary selection specifically for these abilities, since any selection occurred before such abilities were on the scene. Reading

and writing are in large part results of cultural forces that exploit neural coincidences, just as the arts and sports are the result of cultural exploitation of neural coincidences, and many activities would be impossible without some or all of our sensory capacities. For example, 'turn out' in ballet is a very important ability for classical ballet, but it relies on physical characteristics and abilities that evolved quite independently.

We might not think of driving at night or ballet as products of functions, but they do belong in that category since errors can be said to be due to lack of functioning or ill functioning. "Acquired functions" are, then, coincidental abilities exploited by culture to produce new products, some of which do enhance survival and reproduction. Thus there are impaired functionings in reading, writing, ballet and night vision, even though it is not plausible to say that evolution selected for any of these functions.

Thus we can conclude that "well functioning" can apply in some cases where evolutionary selection does not seem to apply. A defense against this conclusion has been proposed by Millikan. Millikan labels acquired functions as derivative functions, and takes them to in effect inherit the trait of being selected for. This proposal appears to be dependent on fallacious reasoning. "Cause" as ordinary understood is transitive; if A causes B and B causes C, then A causes C. "Selected for" is different. If A selected for B and B causes C, then it does not follow that A selected for C. If one selects a fruit and the fruit causes an allergic reaction, then it does not follow that one selected the allergic reaction.

To summarize: driving at night, ballet dancing and reading and writing can all be done well or bad. When they are not done well, we are not functioning in that domain as we should. The cause of such malfunctioning may be due to some component not behaving as it was evolutionarily selected to do, or it may be due to something going wrong with the training that allows the performances we value. Or indeed something else, perhaps in the environment. But we may not be able to tell whether the error lies in the inherited characteristics, in the training or in the environment. And this is a powerful reason for resisting the idea that well functioning in cognitive neuroscience just means functioning as evolutionarily selected to do.

Conclusion

We may choose something, such as an activity, that has short term rewards but possible long term bad effects. Cancer treatment for juvenile diseases may leave the child with a propensity for adult cancer. Some sports can allow one a glorious youth, but a more painful old age than one might otherwise have had. Cartilage, we could say, similarly is a great material in one's youth, but its use in our joints can have downsides. One major problem is that it very often wears out in old age.

An orthopedic surgeon who says of a 65 year old woman who is walking with pain that her knee is not functioning probably means that it is not functioning well. But here evolutionary selection plays no easy role. What happens after reproductive years is a

matter with which evolution can have little involvement. Indeed, we might say that decay, and eventual decline in function, is natural and inevitable.

At this juncture the social constructionists can reenter the discussion and point out that what we count as functioning well can involve what we want in fairly arbitrary way. Let us start off with the small cases. It can seem that we evaluate human functioning as functioning well when it is functioning normally. For example, a baby is developing well when it is developing normally, and we have tables and graphs explaining what that is at each stage. Nonetheless, there can be disagreements about whether someone is functioning well in some domain even when they test as normal. IQ is a case in point; one could decide that functioning well as far as intelligence goes must be more than having the average IQ.

The argument heats up if we turn to issues like sex or race. There are traits we associate with a specific sex. For example, men are aggressive and women are gentle, it is said. Nonetheless, what we might think of as such sex-link characteristics are not distributed among human beings strictly according to sex. We probably all know of women who are aggressive and men who are gentle. However, a boy who is gentle and eschews contact sports may be judged by a wide array of people as seriously defective. He is not developing properly.

Disability theorists have recently argued strenuously that many disabilities are created by a person's society and its demands. Neurally atypical people may be unable, for example, to keep up with the very quick interactions of the neurally typical. But that should not mean that that person is sick or diseased.

What should we say, then, of BPD? To be con't.

Anne J. Jacobson
Somerville College
anne.jacobson@some.ox.ac.uk

Bibliography

- Berg, J., John Dickhaut, Kevin McCabe. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior*, 10, 122-142.
- Carol W Berman, M. (2014). 9 Tips on How to Recognize Someone With Borderline Personality Disorder.
- Davies, Will (2016). Externalist Psychiatry. *Analysis*, 290-6.
- Dennett, D. C. (1969). *Content and consciousness*. London, New York,: Routledge & K. Paul; Humanities P.

- Drayson, Z. (2014). The Personal/Subpersonal Distinction. *Philosophy Compass*, 9(5), 338-346.
- Friedel, R. O. (2004). *Borderline Personality Disorder Demystified*: Marlowe & Co
- Hornsby, J. (2000). Title Personal and Sub-Personal: A Defence of Dennett's Early Distinction. *Source Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, vol. 3, no. 1, pp.
- Kim, J. (2005). *Physicalism, or something near enough*. Princeton, N.J.: Princeton University Press.
- Kim, J. (2007). Causation and Mental Causation. In B. P. McLaughlin & J. D. Cohen (Eds.), *Contemporary debates in philosophy of mind*. Malden, MA: Blackwell Pub.
- King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., & Montague, P. R. (2008). The Rupture and Repair of Cooperation in Borderline Personality Disorder. *Science*, 321(5890), 806-810. doi:10.1126/science.1156902
- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to Know You: Reputation and Trust in a Two-Person Economic Exchange. *Science*, 308(5718), 78.
- Machery, E. (2009). *Doing without concepts*. Oxford ; New York: Oxford University Press.
- McLaughlin, B. P., & Cohen, J. D. (2007). *Contemporary debates in philosophy of mind*. Malden, MA: Blackwell Pub.
- Montague, P. R., & Quartz, S. R. (1999). Computational Approaches To Neural Reward And Development. *Mental Retardation And Developmental Disabilities Research Reviews* 5, 86-99.
- Murphy, G. L. (2002). *The big book of concepts*. Cambridge, Mass.: MIT Press.
- Polger, T. W. (2016). *The multiple realization book* (1st edition. ed.). New York, NY: Oxford University Press.
- Potter, N. N. (2009). *Mapping the Edges and the In-between: A critical analysis of Borderline Personality Disorder*: OUP Oxford.
- Thornton, T. (2013). Clinical judgement and tacit knowledge. In K. B. e. a. Fulford (Ed.), *Oxford Handbook of Philosophy and Psychiatry* (pp. 1047-1062): Oxford: Oxford University Press.
- Wagner, J., & Brooks, R. (1977). *The bunyip of Berkeley's Creek* (1st American ed.). Scarsdale, N.Y.: Bradbury Press.
- Wedgwood, R. (2006). The Internal and External Components of Cognition. In R. J. Stainton (Ed.), *Contemporary Debates in Cognitive Science* (pp. 307-325): Blackwell Publishing Company.
- Wirth-Cauchon, J. (2001). *Women and borderline personality disorder : symptoms and stories*. New Brunswick, NJ: Rutgers University Press.