Sculpting the space of actions: explaining human action by integrating intentions and mechanisms

Keestra, M.

2 CONCEPTS AS DELINEATIONS FOR EMPIRICAL CONTENT[15]

In the previous section, a comparison was made between amateur and expert singers. As we saw, differences were partly due to training and education, which had a differential impact on neural organization of the brain of these two groups and on relevant cognitive and motor processes. Nevertheless, it seems reasonable to include both groups of singers in a cognitive neuroscientific study of singing. In other words, it does not seem sensible to separate these groups and to argue that amateur and expert singers are in fact doing something completely different when they sing – making a comparison between the two groups unacceptable. Moreover, as a determined amateur singer can usually develop into an expert singer after appropriate training, the two are distinct in a gradual sense only. Apparently, no distinction in the underlying processes is enough to dissuade us from treating amateurs and experts alike as objects for a study in singing. However, it is not always so easy to decide whether two distinguishable groups can be considered to be performing the same cognitive or behavioral task.

Sometimes it is difficult to judge if observable differences between subjects force us to split a group into two –or even more- different groups with respect to a particular task. For example, are the vocalizations of monkeys to be considered as singing and can we compare their performance and underlying processes with those of human singers? Or at what moment during child development do we accept a child to be singing and not just making vocalizations? And how about those animals most kindred to us with respect to music: the birds? It may prove difficult to deny that birds are singing, even though there are differences in human and bird song.[16] Are we therefore allowed to compare their cognitive and behavioral processes with those of human singers, or will that not inform us about human singing because of the differences between the two species?

It is such conceptual questions that motivates the methodological approach to cognitive neuroscience advocated in the joint work of neuroscientist Bennett and philosopher Hacker. With their much debated book 'Philosophical Foundations of

[15] The present discussion of Bennet & Hacker's work elaborates on the critical articles that were published together with Stephen Cowley. Our critical review article (Keestra and Cowley 2009) received a rather harsh response in (Hacker and Bennett 2011) which we rebutted in our (Keestra and Cowley 2011). Thanks are due to Stephen Cowley for this collaboration.

[16] Distinctions between human and bird song are often made with reference to their structural properties. Especially in the light of structural properties like syntax and recursivity, qualitative differences between human and bird song seem obvious. However, these distinction then rely on the assumption that singing is for both species a form of communication of meaning – cf. (Hauser, Chomsky et al. 2002).

Neuroscience' (Bennett and Hacker 2003) they make a strong plea to give priority to conceptual analysis of psychological functions under study and to subordinate empirical studies to the a priori concepts of such functions. Their strict distinction between the results of conceptual analysis and scientific research leaves limited room for influences of empirical research on concept definitions. Given the extremity of their position, it provides a useful starting point for our current search for a proper method to align investigations of subjects that perform comparable actions yet in remarkably different ways.

2.1 Concepts as 'clean instruments' for neuroscience

Being a neuroscientist and a philosopher respectively, Bennett and Hacker start their jointly written volume by declaring that: "[i]t is concerned with the conceptual foundations of cognitive neuroscience – foundations constituted by the structural relationships among the psychological concepts involved in investigations into the neural underpinnings of human cognitive, affective and volitional capacities. Investigating logical relations among concepts is a philosophical task. Guiding that investigation down pathways that will illuminate brain research is a neuroscientific one. Hence our joint venture" (Bennett and Hacker 2003 1). After they declare a strict distinction between philosophical and neuroscientific tasks and state their view that there are conceptual foundations involved in neuroscience, we learn that they were motivated by a serious dissatisfaction with neuroscientific writings with regard to these foundations. For they held "a suspicion that in some cases concepts were misconstrued, or misapplied, or stretched beyond their defining conditions of application" (Bennett and Hacker 2003 1). Apart from the question what 'defining conditions of application' imply and what role these have – to which we'll return later – the picture that emerges is that the investigation of concepts does not belong to neuroscience's tasks. On the contrary, neuroscience has to accept and correctly apply the concepts when carrying out its own task. What then is that task, if it is not in any sense involved in the investigation of concepts, or in the construction or the development of new forms of application of concepts?

As we can expect from the above, neuroscience is said to deal solely with empirical issues, as: "[i]t is its business to establish matters of fact concerning neural structures and operations" or to "explain the neural conditions that make perceptual, cognitive, cogitative, affective and volitional functions possible" (Bennett and Hacker 2003 1). Establishing facts and explaining conditions are indeed empirical scientific tasks, but still their being logically distinct from the philosophical task needs to be specified. This is done by means of a parallel: "we distinguish between the statement of a

measure and the statement of a measurement" (Bennett 2007 129) – neuroscientists taking the measure for granted and employing it in their business of measuring their objects. Clarifying the logical difference, the authors go on to say that the statement of measure is 'normative (and constitutive)', while the statement of measurement is purely 'descriptive' (Bennett 2007 130). What does this relation between statement types amount to in neuroscience?

The task of neuroscience can allegedly only take place once the philosophical task of concept analysis has already being carried out. And this task is allegedly not empirical in nature but prerequisite to it. As such, Bennett and Hacker do at times compare the relation between the two tasks with the relation between mathematics and physics, for instance when they write: "[n]onempirical propositions, whether they are propositions of logic, mathematics, or straightforward conceptual truths, can be neither confirmed nor infirmed by empirical discoveries or theories. Conceptual truths delineate the logical space within which facts are located. They determine what makes sense. Consequently facts can neither confirm nor conflict with them" (Bennett 2007 129).[17] The middle sentence captures the nature of the relation between the two tasks: first, a conceptual space must be defined in which, second, empirical facts can be placed. Without a given conceptual space, it seems, empirical facts cannot make sense at all. How would this work?

How would a cognitive neuroscientific study of a particular function like action or consciousness depend upon there being a preliminary conceptual space in which facts about that function have to find their place? Such a study often requires the issues like those mentioned earlier regarding singing to be resolved: can we compare animals and humans, is there a relevant difference between children and adults, and so on. If scientists are investigating consciousness, the authors argue that similar questions can be answered once the logical space is already determined by the apriori, conceptual truths concerning consciousness: "Philosophy is concerned with elucidating the defining features of consciousness (its a priori nature). […] Neuroscience, presupposing the concept of consciousness as given, has the task of investigating the empirical nature of consciousness […]"(Bennett and Hacker 2003 403). Obviously, neuroscience has nothing to contribute to the definitory work, on the

[17] Acknowledging that empirical scientists are not always happy with this division of labour and the immunity from empirical critique that it renders to philosophical analysis, the authors insist upon the non-empirical nature of it and the analogy with mathematics: "[f]or neuroscientists such as Edelman to deplore the methods of philosophers as hopelessly a priori is as misguided as it would be for physicists to deplore the methods of mathematicians as a priori" (Bennett and Hacker 2003 402, cf. pp. 7, 385). What they overlook, however, is that the allegedly non-empirical nature of mathematical theorems is itself disputed in the theory of mathematics (Crowe 1988 ; Lakatos 1976).

contrary. Elsewhere they elucidate their idea of defining an object with the example of a vixen: "an animal can be said to be a vixen if and only if it is a female fox" (Bennett 2007 ; Hacker and Bennett 2011). The example certifies that this definition of a vixen does not include biological or genetic information,[18] but instead remains within the verbal realm. However, the question is whether such a conventional or nominal[19] definition adequately captures the difficulty of defining consciousness or other psychological functions. If defining such functions is more problematic, as we believe it is, this seriously undermines this methodological proposal

Let us explain our doubts with the example of consciousness. Hacker and Bennett responded to our critique of their approach in (Keestra and Cowley 2009) with the acknowledgment of assuming the following: "We *took it for granted* that we all know how to use the word 'conscious' and its cognates –for that is all that is necessary for the clarification of *the concept of consciousness*" (Hacker & Bennett, 2011, p. 411, italics in original). Crucial here is their relying upon a 'we' that 'all know' how to use this word. That their approach deserves to be called 'anti-empirical conceptual analysis' (Sytsma, 2010) is not difficult to demonstrate in the context of consciousness. A succinct survey of philosophical accounts of consciousness shows that competent philosophers have not yet been able to settle their debates concerning consciousness and conscious states (Kriegel 2006), and the presence of heated public debates about animal consciousness and euthanasia of patients in a vegitative state confirms that a public community of competent speakers has not yet universally accepted a particular meaning of those intricate concepts.

Still, according to this proposal, the definition of the concept for a function rests not just upon a single but upon two interdependent sorts of information. First, a definition of a concept relies upon its relation to other concepts. Second, and integral to the meaning of a concept, are the criteria for the use of such a concept in this view. Let us first elucidate the role of conceptual relations. The clarification of concepts and conceptual networks that we use when describing facts is carried out in analytic

---

[18] Defining a fox and even defining femininity can be harder than is often assumed – even though most people would agree on some standard defining features of gender, whereas there may be more instances when doubt about the genus of a given cat-like animal arises.

[19] Even though the authors praise Aristotle for paving the way for their type of criticism of conceptual flaws, they did not recognize that in fact, in Aristotle's Posterior Analytics, there is a transition without strict separation from nominal to explanatory or causal definitions (Charles 2000; Demoss and Devereux 1988) – in contrast to the logical distinction made by Bennett and Hacker. In addition, for Aristotle, explanatory pluralism renders definition of biological functions and properties unlike definition of mathematical objects (Gotthelf 1997). Like Aristotle, I think that this also holds for psychological functions: these vary both in different kinds and within a single individual. Thus bodily aspects are needed in analysis, description and explanation of psychological functions (van der Eijk 1997).

philosophy in the form of a *'description of our conceptual scheme'* (Bennett and Hacker 2003 439, italics in original). This conceptual scheme is nothing new, they themselves say, it is even "the ordinary conceptual framework". Ordinary indeed, for: "[i]t consists of the familiar array of concepts we have all acquired in the course of mastering the humdrum psychological vocabulary of sensation and perception, cognition and cogitation, imagination and emotion, volition and voluntary action, which we employ in our daily lives" (Bennett and Hacker 2003 114). Nonetheless, these ordinary concepts are being compared with instruments, which tend to have a more specific function and use.

Even though the concepts making up the framework are not specifically designed by or for neuroscience, they are said to function inevitably as "spectacles through which psychological phenomena are viewed and understood." Since spectacles interfere with a person's vision, there is a risk involved: "[i]f these spectacles are askew, then neuroscientists cannot but see the phenomena awry" (Bennett and Hacker 2003 115). Apparently, even though spectacles are usually made with a specific function to fulfill or to compensate for a specific person's vision deficit, the authors hold that ordinary concepts can similarly be considered to be askew or not. Confirming this is their statement that words: "are the instruments of thought and reasoning" and their insistence that it "behoves us to be aware of our instruments and to ensure that they are clean" (Bennett and Hacker 2003 381). In sum, in spite of its being ordinary and non-scientific in nature, our conceptual scheme or framework can allegedly be analyzed – and corrected, if necessary - in such a way that it provides lay-persons and neuroscientists alike with correct spectacles or clean instruments. Even thought we doubt the appropriateness of this comparison of concepts with functional instruments, in the next section we will show where the authors believe that we find our conceptual instruments or how we can adjust our conceptual spectacles.

2.2 Connective analysis and ascription criteria

Cleaning our concepts, which we need as instruments, is partly carried out by a method Bennett and Hacker write about in a methodological section on 'Connective analysis in philosophy'. There they write that such a connective analysis: "traces, as far as is necessary for the purposes of clarification and for the solution or dissolution of the problems and puzzles at hand, the ramifying logico-grammatical web of connections between the problematic concept and adjacent ones" (Bennett and Hacker 2003 400). The web of connections should inform about the "logical possibilities" or the "combinations of words [that] are significant and can be used, within or without science, to say something true or false" (Bennett and Hacker 2003

401). The description of this result cannot be compared to cleaning instruments or correcting vision, as the latter activities allow gradual improvement, while logical possibility does not. Indeed, a logical possibility implies a definitive answer to a question like: "[w]hat kinds of things can be coloured – that is, what are *intelligible* subjects of colour predicates" (Bennett and Hacker 2003 130, italics in original). But this latter example is quite specific and involves the logico-grammatical relation between subject and predicate that is of particular concern to the authors and which they discuss with regard to the mereological fallacy that they find to be commonly made in cognitive neuroscience writing. We will come back to that later, but will first consider more closely how a connective analysis can deliver the 'defining features' for a psychological function like consciousness.

The nature of the connective analysis that should deliver the necessary web of connections is rendered relatively clearly at the beginning of the section on one of many forms of consciousness: transitive consciousness. "Transitive consciousness lies at the confluence of the concepts of *knowledge, realization* (i.e. one specific form that acquisition of knowledge may take), *receptivity* (as opposed to achievement) of knowledge, and *attention* caught and held, or given. The various categories or kinds of transitive consciousness that we have distinguished are differently related to these. We shall sketch some of the connecting links and some of the conceptual differences between these loose categories" (Bennett and Hacker 2003 253, italics in original). What can be learnt from this statement is that a particular form of consciousness is indeed being analyzed with the use of – 'adjacent' - concepts that are useful for describing or defining transitive consciousness. For instance, transitive consciousness can be described as a form of knowledge about an object, it being a knowledge that is not actively achieved or attained. Instead, the contents of transitive consciousness are, according to this analysis, merely being noticed, realized or one just becomes aware of them (Bennett and Hacker 2003 253). Such establishment of a conceptual framework when defining a concept does seem useful. What remains unclear, however, is what the source of the relevant web of connected concepts is and precisely how they are so sure about the relations between concepts when describing a phenomenon like this.[20] For instance, one could wonder whether previously attained knowledge influences transitive consciousness, heightening the receptivity of a

[20] There are places when Bennett and Hacker are less certain or where they acknowledge that strict delineations are difficult to achieve. An example is emotions. Notwithstanding the remarkable conciseness of the chapter – only 25 pages, in contrast to some 130 for (self-)consciousness – they introduce an uncommon subdivision of affections into emotions, agitations and moods, only to admit later that the: "boundaries between emotion, agitation and mood are not sharp" (Bennett and Hacker 2003 202).

subject as it does for some objects more than for others. If so, should we distinguish different forms of transitive consciousness? How should we decide such cases, where do we find the criteria to decide one way or another?

The authors do not appear to have much doubt about such matters regarding the source of the array of concepts or their applicability, as was evident from the quote above in which they referred to common knowledge about the use of the word 'conscious' and related words. Apparently, their equation of meaning and use – inspired by their interpretation of Wittgenstein - has found an uncomplicated application in the context of the conceptual foundations of neuroscience, even with reference to not undisputed concepts like consciousness. But these disputes will not easily affect the approach of Bennett and Hacker, since they put some conditions in place such that their assumption of consensus is not easily threatened.

The assumed consensus is grounded in the existence of a community of speakers, which – perhaps tacitly – has determined correct and incorrect explanations for verbal meanings. In so doing, words within such a community have a rule-governed use, which in turn determines their meaning (Bennett and Hacker 2003 382). Two additional conditions further restrict the source of word use grounding the investigated conceptual definitions when the authors state that they rely on "what competent speakers, using words correctly, do and do not say" (Bennett and Hacker 2003 400). The conditions of competence and correctness of use do to a large extent overlap or define each other reciprocally: incompetence in language use is observed especially through the incorrect use of verbal expressions, and vice versa. Combined, these conditions here depend again upon the presence of conceptual consensus within a given community. As a result, the authors modestly claim to offer only: "the ordinary conceptual framework properly elucidated" (Bennett and Hacker 2003 114), intending to be uncontroversial and merely "to outline distinctions which are familiar and in constant use" (Bennett and Hacker 2003 117).

Such consensus must be assumed as it also provides the basis for a speaker's competence to develop: "[a] competent speaker is one who has mastered the usage of the common expressions of the language" (Bennett 2007 146). They illustrate the latter with examples that refer to words black, vixen, perambulate, man and ten o'clock. Avoiding discussion here of the potential disagreements on particular instances of these words, even though we believe these are all less complicated than 'conscious', let us end this section with some more information on the criteria for use, since word use plays such an important role in this approach. Indeed, words and concepts will be found to be of importance for the other methodological approaches as well, so the present discussion prepares us for the treatment of those as well.

Closely related to the connective analysis, laying bare the conceptual framework or the web of connections between concepts, there is a second source of information about their meanings. This source is derived from observing cases in which these concepts are or are not being used. According to the authors, psychological concepts such as consciousness, perception and emotions are being used meaningfully only in relation to other human beings.[21] The rules that appear to be determining such application of these concepts are related and complementary to the rules that determine the connections mentioned above: "[t]he criterial grounds for ascribing psychological predicates to another person are conceptually connected with the psychological attribute in question. They are partly constitutive of the meaning of the predicate" (Bennett and Hacker 2003 83). To such ascription criteria of a psychological function belong particularly the behavioral expressions that are connected to that function. In the case of pain, for example, it is pain-behavior like moaning that is relevant: "Pain-behaviour is a criterion – that is, logically good evidence for being in pain", the authors write, and they conclude: "[t]hat such-and-such kinds of behaviour are criteria for the ascription of such-and-such a psychological predicate is partly constitutive of the meaning of the predicate in question" (Bennett and Hacker 2003 82).[22] They emphasize that the fact that behavioral criteria are partly constitutive of a concept's meaning distinguishes these criteria from being mere inductive evidence.

In contrast to such behavioral and non-inductive evidence, neuroscientific investigation of psychological functions like pain or consciousness, aims to produce inductive evidence concerning the brain events associated with such a function. Preliminary to such scientific investigation, the authors argue, a non-inductive and logically sound ascription of pain or consciousness can and needs to be made to a subject that is being neuroscientifically investigated. That ascription rests upon the investigators using these words correctly, including controlling whether the behavioral criteria for application of these words are being met. If a subject is

[21] Sytsma justly emphasizes that B&H fail to produce empirical evidence with regard to the words that they subject to connective analysis and has consequently called the method of PFN 'anti-empirical conceptual analysis'. To underline this diagnosis he produced empirical evidence that, contrary to the authors' intuitions, a significant portion of subjects don't hesitate to apply the verb 'calculate' to computers –though B&H reject this as nonsensical (Sytsma, 2010).

[22] Debate about behavioral criteria is likely to emerge, especially with regard to an elusive phenomenon like consciousness. Indeed, an fMRI and clinical study of patients diagnosed with only vegetative consciousness has shown that some patients were able to willfully change their conscious state in such a way that it was detectable with imaging techniques, in the absence of any distinct behavioral responses (Bennett and Hacker 2003 202). However, such an approach has been criticized with reference to the behavioral criteria required by Hacker, as in (Monti, Vanhaudenhuyse et al. 2010) – these criteria would still not be fulfilled with fMRI evidence. An interesting alternative has been proposed, namely to use a brain-computer interface as a way of facilitating behavior to patients without any muscular control (Nachev and Hacker 2010).

not meeting those – non-inductive, behavioral – criteria, the inductive evidence derived from the neuroscientific investigation cannot be correctly correlated to the psychological function that the investigators believe to be scrutinizing. The logical order is such that only if the ascription criteria are met, the empirical evidence can be inductively correlated to the alleged function: "if such inductive evidence conflicts with the normal criteria for the ascription of a psychological predicate, the criterial evidence overrides the inductive correlation" (Bennett and Hacker 2003 83). If applied to an example like vixen, this seems evident: if closer inspection of a particular animal that a scientist calls 'vixen' produces evidence that the animal is in fact a female wolf or that it is a male fox, further evidence about it does not apply to vixens, too. This observation does however merit further specification: evidence about gender specific features or about features that are used to distinguish wolves from foxes may no longer be applicable. Therefore, it may make sense only for such limited examples and in a limited sense to declare that: "[c]onceptual truths delineate the logical space within which facts are located" (Bennett & Hacker, 2007, p. 129). Indeed, we may doubt whether interdependence between scientific facts and conceptual truths can be avoided, for example concerning an animal's gender and its precise species definition.

Generally, natural kind concepts and classifications in the life sciences and behavioral sciences lack the kind of unity and demonstrate much more divergence than can be found in domains with less complex and dynamical phenomena, like chemistry (Dupré 2001). The reason is that phenomena studied by the life and behavioral sciences are generally produced by a much greater and wider range of causes, which simultaneously determine these phenomena. Correspondingly, any attempt at delineating a logical space that consistently and comprehensively encloses only those facts that pertain to an allegedly definable psychological function must take a variety of criteria and logical connections into account. Otherwise, the conceptual space runs the risk of resting upon ill-founded assumptions about a domain's contents, like its definability and the uniqueness of its corresponding definitions (Hacking 1991). In the final section on this approach that aims to define conceptual foundations of neuroscience, we will demonstrate where it runs into trouble and what consequences can be drawn for the relation between psychological functions, the concepts that correspond to these and neuroscientific evidence with regard to these.

2.3 Non-convergent and variable criteria, and their implications

In a field where causal pluralism affects relevant phenomena, exceptional and

surprising cases will likely obtain. Referring to our example of singing again, we doubt whether people would always agree in deciding whether or not a person who is making vocal sounds is singing. At what age do we ascribe 'singing' to an infant and not just babbling? Similarly, are religious recitations instances of singing, or rather peculiar intonated readings? Will we agree on when a speaker of a tonal language, like Mandarin, has shifted from speaking to singing? Are there perhaps even cases in which we ourselves unwittingly made such a shift? The blurred distinctions between speech, recitation, babbling and singing as well as our probable disagreements suggests that such concepts are in fact formed and used as prototypes, rather than definable under the conditions suggested by Bennett & Hacker.[23] Since there are conventional concepts like 'bachelor' or perhaps 'vixen' that are rule-governed, our language probably contains both prototypical and rule-governed concepts (Ashby and Ell 2001). The advantage of concepts as prototypes is that speakers can apply such concepts with some liberty and still remain understandable.[24] A strictly delineable conceptual space does not allow such liberty in use, as the disputed status of 'blind-sight' demonstrates.

This example refers to investigations of a famous patient, who was found in specific behavioral experiments to demonstrate 'good visual discrimination capacity in the absence of acknowledged experience' (Weiskrantz, 1997, p. 19) – behaviorally responding above chance to a stimulus, whereas she explicitly denied perceiving that stimulus. From this surprising combination of behavioral evidence for, yet verbal evidence against perceptual discrimination in this patient, Weiskrantz concluded that she suffered from 'blind-sight' (Weiskrantz 1997 19). This was how he addressed the issue that the facts collected in studying this patient did not permit insertion in either the conceptual space for 'visual perception' nor in the space for 'blind'. Indeed, if we consider these spaces for a moment as 'logico-grammatical' Venn diagrams, one could even imagine that the spaces 'visual perception' and 'blind' overlap at some point. In this overlapping area, then, the facts pertaining to this patient could be located. However, Bennett & Hacker argue otherwise.

---

[23] Stokhof points out that for Wittgenstein it is not just the normative practice of rule-following that constitutes the meaning of concepts. In addition to these, constraints imposed on our practices by our environment and our human nature have an impact on our conceptual schemas (Stokhof 2000), which are not articulated in Bennett & Hacker's approach.

[24] A prototype theory of psychological concepts has been proposed – on various grounds – by Paul Churchland (Churchland 1988). Elaborating on the ideas of Churchland and others, another conception of concepts as state spaces can be found in (Gärdenfors 2004a). These authors contest the assumption that concepts are always rule-governed (or symbolic). Although our approach to the process of 'sculpting the space of actions' has some affinity with theirs, we aim to show how rational considerations can also contribute to this process of sculpting a state space that pertains to actions.

Their conclusion regarding 'blind-sight' is straightforward and relies on their assumption of the strict definability of psychological concepts, partly constituted by their behavioral criteria. To begin with, they observe that in this patient "the normal convergence of indices of sight –namely, appropriate affective response, behavioural reaction, reoriented movement, verbal description, answers to appropriate questions, etc. – is subtly disrupted." Then they refer to their assumption that "such *convergences constitute the framework* within which verbs of vision are taught and used. (...) The consequence of a conflict of criteria is that one can neither say that the patient sees objects within the scotoma nor say that he does not." Finally, their conclusion from this is that this patient's case "indicates the *inapplicability of a concept* under special circumstance" (Bennett & Hacker, 2003 396, italics not in original). With concepts that function as prototypes, this conclusion of conceptual inapplicability is avoidable. Apart from the conceptual dispute, it is important to realize the consequence for the empirical evidence gathered by investigating this patient: according to Bennett & Hacker it will have little relevance for the explanation of normal vision. Before explaining this position and then contrasting it with the cumulating evidence for the divergence with regard to psychological concepts and behavior, let us underline what is at stake in the present case of blindsight.

Given their assumption that psychological concepts can be assigned strictly delineated logical spaces for which both logico-grammatical and behavioral criteria are to be used, there is principally no room for divergences regarding the use of those concepts. This also holds for those cases where some criteria for the use of a particular concept are met, while other criteria appear to be contradicted. Such divergence of criteria would allegedly render a concept meaningless and consequently useless. Accordingly, a concept is never applicable in those situations in which conflicts arise with respect to the criteria that should determine the use and hence the meaning of the concept. In contrast to a prototype theory of concepts that allows some distortion and divergence in the formation and use of concepts (Ashby and Ell 2001), as does a theory of concepts that projects a multi-dimensional state space for a concept (Gärdenfors 2004b), Bennett & Hacker cannot permit any flexibility in the criteria that constitute the meaning of concepts. Indeed, in their response to our critical review article (Keestra and Cowley 2009), they compare the correct application of concepts with following the rules in a game where those rules in fact constitute the game. This odd metaphor brings them to take on a judge-like function when writing: "Far from delimiting neuroscience or narrowing its scope, we constrain it only in the sense in which one constrains draught players in pointing out that there is no checkmate in draughts – which is no constraint" (Hacker and Bennett 2011

461). However, the arguments here and in our rebuttal (Keestra and Cowley 2011) suggest that if this metaphor of concepts as rule-governed games holds at all, it has very little value for psychological concepts. For with regard to psychological concepts we should expect, for various reasons, a variability and divergence that makes a different theory of concepts more appropriate. Such a theory could then also allow the scientific investigation of an extraordinary case like blindsight some relevance for the explanation of normal vision. Bennett & Hacker, on the other hand, cannot allow such an applicability of insights in blindsight to cases of normal visual perception.

The reason they offer to deny that a patient that we diagnose as and call 'blind-sighted' can yield any neuroscientific insight on perception is as follows. Given their assumption that the behavioral criteria are partly constitutive of the concept 'seeing' or 'vision', the acceptance of the contradictory criteria that are applicable to this patient would in fact imply that we change the concept itself. Given the connections between concepts – that are subject of a connective analysis – such a conceptual change could not be made without in turn modifying all those concepts related to 'seeing' or 'vision'. Eventually, the consequences would be wide-ranging for many concepts and phrases in which these figure. When redefining a word like 'perceiving' or 'remembering', neuroscientists would be obliged to do the following: "New formation rules would have to be stipulated, the conditions for the correct application of these innovative phrases would need to be specified, and the logical consequences of their application would have to be spelled out. Of course, if this were done, the constituent words of these phrases would no longer have the same meaning as they have now. So *neuroscientists would not be investigating the neural conditions* of thinking, believing, perceiving and remembering at all, but rather those of something else, which is as yet undefined and undetermined. But this is patently not what neuroscientists wish to do" (Bennett and Hacker 2003 384, italics not in original). Or, applying once more their metaphor mentioned in the previous section, the neuroscientists that investigate 'blind-sighted' patients would play chess while those that focus on normal vision are playing draughts or even baseball – precluding any useful exchanges or competition between the two.[25]

An interesting asymmetry emerges between neuroscientific results pertaining to an exceptional case like a patient with 'blind-sight' and results pertaining to normal

[25] In the terms that Christensen & Sutton use in their discussion of an integrated approach to moral cognition, Bennet & Hacker would assume that it is possible to construct a 'clean taxonomy' for such a cognitive function. Christensen & Sutton, in contrast, argue that we cannot avoid 'messy taxonomies' for such functions due to the "complex underlying causal factors that overlap across categories" (Christensen and Sutton 2012).

subjects. Indeed, although the authors believe that although neuroscientists "can brilliantly explain why patients cannot behave as normal humans can in a multitude of different ways" (Bennett and Hacker 2003 365), explaining normal functioning cannot refer to such neural conditions. In contrast with explanations of pathological behavior, to "explain typical human behaviour, one must operate at the higher, irreducible level of normal descriptions of human actions and their various forms of explanation and justification in terms of reasons and motives (as well as causes)" (Bennett and Hacker 2003 365). Even though causes are added – albeit in brackets – to the list, these are not subsequently clarified like the other ingredients. So it remains unclear whether these causes refer to the tendencies or habits of an individual, to the moral and social norms or to other not explicitly mentioned ingredients. In any case, the authors appear to render only a secondary role to causal conditions when humans are explaining each other's typical behavior, even though humans typically accept that causal conditions at the neurophysiological level do offer bottom-up constraints on someone's behavior. We will come back to that in section I.2.4. Here, we would like to add another argument why we believe that the authors' assumptions are not warranted, suggesting as they do that it is always possible to make a clear distinction between normalcy and pathology and suggesting that there is always consensus concerning the use of psychological concepts within a community of competent speakers.

In contrast to these assumptions, researchers tend to accept that divergence is prevalent in the context of behavioral criteria, concept use and even neural correlates of human psychological functions. Textual analysis and interpretation have long suggested historical and cultural diversity in these contexts (Lloyd 2007 ; Snell 1975). In addition, psychological and psychiatric experiences suggest that subjects of different cultures do not only differ in the use of psychological concepts but also in their expectations of behavior corresponding to the psychological functions referred to with these concepts (Chaturvedi and Bhugra 2007).[26] Additional insights on etiology and clinical phenomena support the proposal that the strict distinction between pathological and normal states cannot be upheld, whereas a more gradual distinction between those states seems more plausible (cf. (Hyman 2007 ; Newsome, Scheibel et al. 2010)), adding to the divergences.[27] A main reason that such divergences

[26] Arguing for a more dynamical mode of classification, Hacking points to fact of a 'looping effect' of psychological and psychiatric classifications. Such classifications tend to influence the groups to which they apply, making people labeled as ADHD or multiple personality disorder patients behave according to the criteria currently used by a classification system like the DSM (Hacking 1995).

[27] Hyman, a member of the DSM-5 Task Force, is highly critical of the classificatory 'silos' of the current and future editions of the DSM. One of the arguments against its classification is that it corresponds poorly with clinical and scientific evidence about distinctions. He suggests integrating clusters of interrelated syndromes into larger clusters – avoiding the assumptions of strict borders between diagnoses altogether and allowing room for additional scientific insights in this context (Hyman 2011).

may have escaped notice of scientists from various fields is that the overwhelming majority of subjects used in research are drawn from a very small and specific selection of the world's population (Arnett 2008 ; Henrich, Heine et al. 2010). As emerging results of transcultural neuroscience show that transcultural differences are likely to affect not just functional networks but perhaps even anatomical structures in the brain (Han and Northoff 2008), this limitation of research subjects has serious implications for the validity and significance of its results. Such divergences are due to the long-time exposure to different cultural experiences and behavioral practices (Park and Huang 2010). Given such evidence and accounts of divergences in neural activations and structures, in behavioral experiences and criteria, and in psychological concepts, there is reason to question the consensus within a community of competent speakers, as assumed by the authors' approach. If this consensus is to be found both in concept use and regarding behavioral criteria, it may be limited to a rather restricted community. Although the authors' ambitions are larger, their conceptual foundations of neuroscience may in fact not transcend its origin as a form of "contemporary English philosophical anthropology" (Quante 2008), as a reviewer of Hacker's categorical account of human nature has elsewhere suggested. Avoiding such a serious limitation, in the next and final section on this approach, we will defend a more liberal stance with respect to conceptual and behavioral divergences, while sustaining Bennett & Hacker's critique on mereological fallacies in neuroscience.

2.4 Heuristic use of conceptual divergences, yet with limitations

In section I.2.1 we found that the project of developing conceptual foundations of neuroscience was mainly inspired by "a suspicion that in some cases concepts were misconstrued, or misapplied, or stretched beyond their defining conditions of application" (Bennett and Hacker 2003 1). Neuroscientists do tend to offer factual results of neuroscientific investigations as having implications for our interpretation of psychological concepts. That is, they sometimes believe they can redraw a conceptual space on the basis of those facts, instead of merely gathering facts that either belong or do not belong to a particular, predefined space. This alleged neuroscientific hubris and misapplication is warded off by presenting a 'mereological principle', which in itself is a consequence of the authors' analysis of the nature and origin of psychological concepts, as outlined above. The principle states that: "psychological predicates which apply only to human beings (or other animals) as wholes cannot intelligibly be applied to their parts, such as the brain" (Bennett and Hacker 2003 73).[28] Doing so, Bennett & Hacker argued, would be similar to chess players applying the rules for draught or bridge and thus constructing an altogether different game.

The motivation for the mereological principle depends partly on their rejection of ontological and explanatory reductionism. Scientific reductionism, they write, "is a commitment to the complete explanation of the nature and behaviour of entities of a given type in terms of the nature and behaviour of its constituents" (Bennett and Hacker 2003 357). In the case of neuroscientific analysis of human cognition and behavior, reductionism would imply that these are completely explainable in terms of neurons and neuronal activities. Bennett & Hacker have warded off this threat of reductionism by strictly separating the analysis of psychological concepts logically from the collection of empirical facts and, second, by disallowing the application of those concepts to objects other than the persons to which competent speakers ascribe them. This leaves no room for any identification of cognition and behavior with the properties, activities or interactions of neurons. However, as we will argue in this section, they overlook the possibility that other relations between psychological concepts and the study of the relevant neurons or neuronal activities are possible and even fruitful. Indeed, we will argue that a challenge for cognitive neuroscience is to develop a more useful integration of conceptual analysis and empirical research.

For Bennett & Hacker, it is straightforward that on the basis of our knowledge of the conceptual scheme of psychological concepts and of our observation of a person's behavior that we can only conclude that this person is perceiving or knowing or feeling – and not in any sense that his brain or neural areas are performing those functions. Obviously, the brain and neural areas are involved in producing a person's behavior but only in the sense of: "causally necessary conditions for the human being to think or perceive, imagine or intend" (Bennett and Hacker 2003 117). Given the nature of the sources of our conceptual truths, there is no room in this approach for these causal conditions to be more directly related to concepts like thinking, perceiving, imaging or intending– or something like their 'concept spaces'. Instead, causal conditions or correlates, being the result of empirical research, are held to be logically different and separate from those conceptual truths.

If, however, our arguments above are sound, then this strict distinction and the endeavor as a whole is flawed. If, that is, the assumption of consensus regarding

[28] A critique of the authors' limited account of mereological reasoning, their overlooking of the heuristic use of such reasoning and their misinterpretation of Aristotle's warnings in this context was given in (Keestra and Cowley 2009). Though largely dismissing our critique in their response, they did not address this issue (Hacker and Bennett 2011). We, in turn, reconfirmed our limited agreement with their mereological principle, albeit for different reasons (Keestra and Cowley 2011). The relativism inherent in Aristotle's analysis of part-whole relations is also commented upon in (Koslicki 2007). This relativism is better accounted for in the mechanistic explanatory approach, which is also interested in constitutive relations yet explicitly acknowledges the validity of an explanatory mechanism for a particular phenomenon.

psychological concepts within a community of competent speakers is unwarranted and if a consistent and comprehensive delineation of spaces for such concepts is illusory, then we must look for a different relation between empirical, neuroscientific facts and conceptual insights. Consequently, concepts can be considered differently and may yield insights different from those allowed by the approach of Bennett & Hacker. As we will see, the other methodological propositions that we will be discussing in this part suggest a different relation and do allow different roles for concepts and conceptual analysis. Let us finish here by discussing a few possible implications of understanding this relation differently.

If it is impossible to provide a comprehensive and consistent delineation of conceptual spaces pertaining to psychological functions, then we may need to accept and even explore the conceptual divergences and uncertainties that abound in this domain. For instance, competent language users commonly refer to the phenomenon of distraction of attention from pain. Admitting that this phenomenon defies their assurance that "there is no difference between having a sensation and feeling a sensation", Bennett & Hacker refer to this phenomenon as a "curious anomaly" which "can be viewed as a singularity (in the mathematical sense) in the grammar of sensation" (Bennett and Hacker 2003 121, footnote 2). What they fail to do, however, is to take the expression seriously – even though it blurs some alleged conceptual distinctions – and to explore its value as a heuristic. Such a heuristic use of a concept that is hard to define can point us in the direction of an explanation of its intricate character (Keestra and Cowley 2011).[29]

For instance, it may be that the causal conditions involved in pain and in attention do interfere at times with each other, producing this curious phenomenon – as was shown to be the case (Valet, Sprenger et al. 2004).[30] Indeed, when such a phenomenon is being explained with reference to a complex and dynamic explanatory mechanism – which will be clarified more generally further below in this part – its exceptional nature can be ascribed to uncommon interference of components or operations, or to external conditions that influence the explanatory mechanism such that it produces an irregular behavior. Consequently, the apparently strange concept use then

[29] Another type of 'bi-directional' interactions between conceptual analysis and empirical research is presented in Northoff's neurophilosophical methodology (Northoff 2004). Kindred as his approach is, it involves a particular use of philosophical analysis and pays not so much attention to the heuristic use of conceptual divergencies, for example.

[30] A similar explanation has been offered for synaesthetic experiences, which appears to correlate with cortical hyperconnectivity (Rouw and Scholte 2007). B&H pointed out that it makes no sense to ascribe colour to numbers (Bennett and Hacker 2003), which from a strictly semantic point of view may be correct but denies such a concept the role of a heuristic for further investigation of an exceptional psychological phenomenon.

correlates with the exceptional behavior of an explanatory mechanism that produces a surprising phenomenon.

In the case of 'blind-sight', a similar implication may be drawn. It is hard to define comprehensively both 'seeing' and 'being blind', even in healthy subjects, as odd perceptual phenomena occur which suggest temporary or specific forms of blindness.[31] The concept 'blind-sight' signals this blurred and porous character of conceptual definitions. Moreover, it also captures the divergence that is generally observable in the realm of psychological functions, even though we might agree in the case of 'blind-sight' that it refers to an exceptional and pathological phenomenon. Because of divergence and the corresponding ambiguity of psychological concepts in normal situations, competent speakers may at times refer to explanatory components in order to disambiguate their concepts. Given the variety of explanations, such explanatory components may be of various natures.

Aristotle, for example, aimed to define anger by including both a psychological and a physiological explanatory component in it, when he referred to anger as requiring a definition: "as a certain mode of movement of such and such a body (or part or faculty of a body) by this or that cause and for this or that end" (De Anima 403 a 27-28).[32] Filling in the required causal pluralism involved in such a definition, he specified anger as being both "a craving for retaliation" and "a surging of the blood and heat round the heart" (De Anima, 403 a 31- b 1). More generally, Aristotle accepts that such a causal pluralism is involved in human behavior and cognition, including nature (Murphy 2002). In the previous section, we adduced arguments that confirm this causal pluralism to be effective in causing divergence and corresponding conceptual ambiguities or misunderstandings. Further below, we will discuss how it is that such causal pluralism can be held responsible for the divergences that obtain in the domain of psychological functions and that transpire to the conceptual scheme when describing or explaining such functions. Instead of holding on to strict conceptual delineations that are illusory, a different handling of psychological concepts seems in order. A critical yet more tolerant conceptual analysis can indeed be more conducive to empirical research. That is, the use of the concepts themselves should be different, and the relation of the concepts to the facts derived from neuroscientific investigations can be established differently.[33] An

[31] Many such phenomena depend on typical perception-action loops, causing Noë to defend an enactive view of perception (Noë 2004).

[32] As we noted in (Keestra and Cowley 2009), Aristotle is not opposed to mereological reasoning, as it can perform useful functions in science. Pellegrin even argues that Aristotle's biology is in fact a mereology, a study of parts (Pellegrin 1987) – which seems to me to neglect the prominence of Aristotle's ambition to integrate the various causal contributions to a function or a kind.

influential methodological proposal which does so, has been made by computational neuroscientist David Marr. It is to this proposal that we will now turn.

[33] At this point we would like to refer to a comparable interdisciplinary endeavor as Bennett & Hacker's, though with a strikingly different tenor. Hermeneutic philosopher Ricoeur and neuroscientist Changeux agree, in contrast to them, that in this domain a fair amount of semantic tolerance is inevitable, if not without semantic criticism. Although acknowledging the risk that when neuroscientists employ 'semantic short-circuits' they are "illegitimately converting correlations into identifications," they do not aim to correct this with the assumption of strict delineations of conceptual spaces (Changeux and Ricoeur 2000 40).