



Universiteit  
Leiden  
The Netherlands

## Understanding existential self-understanding : philosophy meets cognitive neuroscience

Stee, M.J. van

### Citation

Stee, M. J. van. (2017, June 21). *Understanding existential self-understanding : philosophy meets cognitive neuroscience*. Retrieved from <https://hdl.handle.net/1887/49803>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/49803>

**Note:** To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/49803> holds various files of this Leiden University dissertation.

**Author:** Stee, M.J. van

**Title:** Understanding existential self-understanding : philosophy meets cognitive neuroscience

**Issue Date:** 2017-06-21

## 4

### **The Future of CNS and Contributions by Philosophy**

Cognitive neuroscience (CNS) is a young research discipline. Part of its appeal lies in its promise. In this chapter, I investigate what a future, conceptually improved version of CNS could contribute to human self-understanding. I propose several conceptual improvements for CNS and show the contributions philosophy could make to them.

We saw in chapters 1 and 2 that current CNS research into love and self-reflection does not contribute to our understanding of existential self-understanding. Although CNS develops at a rapid pace, we saw that the CNS research process contains crucial moments of interpretation that are not dealt with critically. The same tasks are employed over and over again. When it comes to interpreting results, researchers often copy interpretations by others instead of examining early interpretations critically and developing alternatives. In chapter 3 we looked at some of the contributions philosophy makes to our understanding of existential self-understanding. In the process, we also saw how philosophers deal with interpretation in ways that promote the quality of interpretations. In this chapter I investigate what future CNS may have to contribute to our understanding of the human condition. I aim to contribute to the improvement and extension of CNS' research by applying some of the lessons to be gleaned from philosophy to the CNS research process.

Others also have argued that philosophers and cognitive neuroscientists should cooperate with one another. Ever since Patricia Churchland's *Neurophilosophy* (1986), people have been trying to do 'neurophilosophy' and 'philosophy of neuroscience' and more recently also 'neuroethics', both in its guise of 'ethics of neuroscience' and of 'neuroscience of ethics' (Roskies 2002). Generally speaking, philosophy of neuroscience and ethics of neuroscience use cognitive neuroscience as a case study to be examined with the toolkit of questions and methods philosophers always use. 'What constitutes a good explanation?' is a question a philosopher of science might ask and similarly, a philosopher of neuroscience could ask 'What constitutes a good explanation in cognitive neuroscience?' Likewise, ethicists of neuroscience may wonder, as other applied ethicists do for other cases, whether particular research techniques used in CNS pose ethical problems. Researchers working in neurophilosophy or the neuroscience of ethics on the other hand engage in naturalized philosophy: they take findings from CNS to

speak to longstanding questions about, for example, the relation between mind and body, or the nature of important concepts underlying morality, such as free will or personhood. Throughout the bridge-building endeavors there is much talk of a co-evolutionary method (a term introduced by Churchland), of the need for interdisciplinary conversation, for collaborative efforts between philosophers and scientists (Churchland 1986; Gallagher 2000; Gallagher 2013; Roskies 2002; Levy 2011). Most of the initiatives just mentioned pursue philosophical questions, however, drawing on cognitive neuroscience. The results of research feed back into philosophy, not into neuroscience. Philosophical theories are expanded or adapted, whereas cognitive neuroscientists continue business as usual.

Another type of bridge-building does aim to affect how CNS research is performed and draws on a philosophical tradition called phenomenology. It has especially been employed in consciousness studies. When the aim of a CNS study is to investigate the neural correlates to conscious experiences, it is argued, CNS researchers might as well make use of phenomenological analyses of the structure of conscious experiences to ensure they understand their object of study and develop proper tasks. Philosophers have outlined the phenomenology they deem relevant for cognitive (neuro)science, noting parallels between the two enterprises (Gallagher and Zahavi 2008; Zahavi 2005). They leave it up to others to figure out how the phenomenological analysis could be implemented into research exactly. Several CNS researchers have taken up the challenge, employing a neurophenomenological paradigm (Varela 1996) to study the neural correlates of attentive states (Lutz et al. 2002) or front-loading phenomenological analyses into experimental designs (Gallagher 2003) to study the neural correlates of the phenomenological distinction between sense of ownership and sense of agency (e.g. Farrer and Frith 2002). In practice, such studies are mostly performed by neuroscientists who also have a background in philosophy or who closely collaborate with philosophers.

Then there are philosophers who aim to further CNS research through criticizing concepts and lines of reasoning that occur throughout CNS as a whole. Examples are critiques of 'localization' (Uttal 2001) or the 'mereological fallacy' of ascribing psychological attributes to the brain instead of the person: the brain thinks, the brain doubts, the brain loves, etcetera (Bennett and Hacker 2003; Bennett et al. 2007). The constructive contributions these philosophers make to CNS are often less clear-cut than their critiques. Maxwell Bennett and Peter Hacker try to provide overviews of how problematic concepts should be used instead and state that they "hope that these conceptual overviews will assist neuroscientists in their reflections antecedent to the design of experiments" (2007, 13). For reasons that will become clear throughout the course of this chapter, I very much doubt that dropping off concepts at the doorstep of

cognitive neuroscientists, as it were, is likely to be useful for setting up CNS experiments. As with bridge-building between CNS and phenomenology, one needs closer collaboration between philosophers and neuroscientists than that.

This chapter consists of three parts. First, I analyze the strategies that CNS researchers currently use in developing their tasks. When trying to improve on the current situation, it is important to be aware of the motivations that drive researchers and the problems that result from current task development strategies. In various ways, task validity is at the heart of the matter. Few CNS researchers are aware of these issues; those who are put their hope in databases. Yet databases cannot address validity problems, nor do innovations such as decoding paradigms circumvent them.

In the second part of this chapter, I venture that philosophy can be of help. Neurophilosophy could be philosophy *for* neuroscience and its findings could feed back into CNS, rather than being philosophy *of* neuroscience, whose results are discussed amongst philosophers only or mainly. Most importantly, conceptual review of CNS literature could provide the reflexive moment that is currently lacking in the CNS enterprise. I introduce what I mean by it and how it is to be done. I discuss front-loading phenomenology and critical analysis of basic concepts more briefly and explain how these too would be improved if they were accompanied by conceptual review. Interpretation of results can be improved if cognitive neuroscientists would diversify interpretations and critically reflect on existing ones.

In the final part of this chapter, we assume all these conceptual improvements will be implemented and CNS will develop on other fronts too. What may CNS contribute to our understanding of the human condition—existential and otherwise—in that future? I argue CNS first and foremost has contributions to make in those circumstances in which it is relevant to consider ourselves as embrained beings. Triangulation of neuroscientific data with data obtained from behavioral studies set up by psychologists may also be useful. I illustrate the relevance and some of the limitations of CNS' contributions to self-understanding through the case of the sleepy teenagers. Finally, I draw some conclusions.

## **1. Where current cognitive neuroscience is headed**

The research process in cognitive neuroscience (CNS) contains two major moments that are based on interpretation in a way that has conceptual implications: setting up experiments through developing tasks and recruiting participants; and interpreting the

neural results in terms of what they may have to say about behavior and lived experience. Interpretation during operationalization is the most fundamental of the two, as interpretation of results depends on it. Interpretations of neural results in terms of behavior and experience gain quality as a larger number of studies is compared. The comparability of studies depends crucially on the comparability of the tasks employed. In this section, I address current task development practices and the risks these involve for the validity of CNS research. Decoding and other analysis developments do not circumvent these issues. Databasing is presented as a potential avenue to overcome some of the problems, yet only relocates or even exacerbates them. In the next section, I propose that conceptual review can address issues of validity and also present other ways in which philosophy may aid cognitive neuroscience.

### **Current task development strategies**

Cognitive neuroscientists have to take several factors into account when choosing what task to employ. Some factors have to do with feasibility. In the case of fMRI studies, for example, participants should be able to perform the task whilst lying absolutely still in a narrow tube and being exposed to significant noise. Other factors are of a more principal nature. As in all experimental research, CNS researchers should make sure that their tasks are valid and reliable inducers of the experience or process under study. A task is valid when it measures what it purports to be measuring or induces what it aims to induce (Kelley 1927; Borsboom, Mellenbergh, and van Heerden 2004). A task is reliable when it is consistent in what it produces under repeated uses (e.g. Mueller 2004).

Reliability affects the likelihood of gathering significant data. If a task induces different experiences in different people, it becomes less likely that significant data patterns in the correlated neural activity will be found. The same holds if a task induces different experiences in the same people at different points in time, or if different stimuli used within a task elicit different experiences. It is possible to test the reliability of tasks and pilot studies are mostly used to this end. (More about pilot studies soon.)

Validity is arguably more important than reliability: a task is of little value if it reliably induces an experience or process that is not the experience or process one is interested in (Mueller 2004). Yet, validity cannot be assessed through a pilot study. Checks on validity exist, but these all consist of comparing a task to another task that is already considered to be valid. Yet on what basis do we deem that task to be valid then? In the end, establishing whether a task measures what it purports to be measuring cannot be done empirically. In CNS, task validity is paid very little attention to. There are no incentives to do so and researchers also do not know how to do so systematically.

In practice, reviewers serve as gatekeepers; they should accept a task for a study to be published. Reviewers are neuroscientists themselves, however. They are not trained to reflect on task validity and there is no reflection on it in CNS handbooks and the like, where discussion of research methods always revolves around discussion of research techniques (e.g. Gazzaniga 2014; Senior, Russell, and Gazzaniga 2009).

Even philosophy of cognitive neuroscience does not contain systematic reflection on the validity of tasks. In the reader *Philosophy and the Neurosciences*, only one chapter deals with the importance of research instruments and techniques for research outcomes. This is mostly a summing up of the challenges stemming from studying the brain in various ways (lesion studies, electrophysiological studies, different types of neuroimaging studies, etcetera). At one point, there is a brief remark stating that “[i]maging studies [...] depend critically on construction of tasks for which researchers already have a plausible cognitive decomposition which they employ to guide interpretation of the imaging results. [...] But these decompositions are themselves contested. [...] any imaging study is only as good as the assumption of decomposition of processing components on which it relies” (Bechtel and Stufflebeam 2001, 70). They leave it at that. There is no hint that a systematic way of dealing with these issues is possible, or that philosophers may have something to contribute in that area. Likewise, the *Oxford Handbook of Philosophy and Neuroscience* contains one chapter on methodology only. Its authors state that “Methodologies that reliably produce effects are often used hundreds, even thousands of times by researchers with widely divergent theoretical outlooks” (Chemero and Heyser 2009). The authors do not discuss the problems this may lead to in general, but instead focus on one particular task: rodent object exploration. They show how carelessly researchers deal with setting up the task and how sparse the information is they provide in their research papers on the object that rodents explore. In spite of this being a contribution to the *Oxford Handbook of Philosophy and Neuroscience*, the authors do not give an idea about how philosophy may help alleviate the validity issues that ensue.

In practice, I find, researchers are lead mostly by considerations of feasibility and reliability when choosing which task to employ. This makes pragmatic sense: a study needs to be feasible to be performed at all and reliability contributes to the likelihood of finding significant, thereby publishable, results. Several strategies to develop or choose a task are open to researchers. The first and main one is to repeat what others have done, i.e. to use tasks that were used in previous studies. We saw this in CNS of love and CNS of self-reflection and we read it in the quote from the *Oxford Handbook* just now: a task that is known to produce significant results is used over and over and over again.

Yet when a field of study is new, as CNS of love and CNS of self-reflection were not so long ago, there is no task precedent on which to fall back. What do researchers do in such cases, or in those cases in which they decide they would like to approach their object of study differently? One way is to try and use an existing task, but from an adjacent research literature. The task used in CNS' self-reflection studies, for example, stems from research literature on the self-reference memory effect. 'Self-reference memory effect' refers to the finding that people remember personality adjectives better when they have processed them in relation to themselves than when they have processed adjectives in other terms, such as whether they are written in uppercase or lowercase (Rogers, Kuiper, and Kirker 1977). That is, the task used in CNS of self-reflection was originally developed for psychology experiments on memory (see also Zahavi and Roepstorff 2011). This explains why in early studies, participants had to recall as many adjectives as they could after the brain scanning had taken place. It also explains why some researchers labelled their studies as 'CNS of self-referential processing'. Later, researchers dropped the memory task after the experiment and used labels like 'self' and 'self-reflection' much more than 'self-referential processing'.

Another way to develop a task for a new research field is to think one up yourself and then spend a few lines in your article arguing for the relevance of the task. No clear-cut case of this can be found in either CNS of love or CNS of self-reflection. CNS of love's first study comes closest to employing this strategy. "Visual input plays a leading role in arousing and sustaining romantic love and yet nothing is known of the neural mechanisms underlying this complex sentiment" Andreas Bartels and Semir Zeki write (2000). That is, they employ pictures to induce an experience of love in their participants and argue for the validity of doing so by stating that 'visual input plays a leading role in arousing and sustaining romantic love'. Subsequent studies in CNS of love repeat Bartels and Zeki's operationalization without discussion, let alone argument for their choice. They simply refer to an earlier use of the operationalization, which is argument enough for reviewers.

When developing a novel task and/or novel stimuli, pilot studies are an important tool researchers can employ. In CNS of love, Debra Mashek, Arthur Aron and Helen Fisher (2000) interviewed people who professed to recently having fallen 'madly in love'. They asked them what things remind them of their relationship. Subsequently, subjects had to sort these things according to how intensely positive they made them feel; how often they encountered them; and how important they were to them. Of all the things that were mentioned, Mashek and colleagues chose 5-10 items, based on the sorting scores and on how easily stimuli could be obtained. They thought up control items themselves. For example, participants reported that certain songs strongly induced romantic love in them. Researchers then paired that song with the alphabet

song sung by Sesame Street characters as a control condition. Likewise, reminiscing key moments in the relationship was paired with remembering the most recent teeth brushing session, as an event of low emotional significance. Pilot participants had to return after 7-10 days and rated stimuli for how intensely they felt love throughout being exposed to them. The items for which the contrast between experimental and control condition was the largest, were deemed to be most suitable as stimuli in CNS of love.

It may seem as if pilot studies provide an empirical alternative to interpretation when it comes to task development. Yet what Mashek and colleagues ensure with their pilot study is that the group on whose self-understanding stimuli and tasks are based in CNS research is larger than the researchers alone. They perform a kind of survey of people's ideas about what would elicit a love experience in them. They do not critically question this self-understanding, however. What is more, they come up with control conditions themselves and do not critically reflect on these either. What they test for is the size of the contrast between experimental and control condition, not the validity of the contrast as an operationalization of love experience in itself. In absolute terms, they find that photographs, songs and reminiscing about the beloved induce most intense love experiences. As far as contrasts are concerned, photographs of the beloved vs. photos of unknown others and photos of natural landscapes show the strongest effect. Reminiscing is also in the top three contrasts, as is smell reminiscent of beloved vs. smell of water. In contrast, the Sesame Street alphabet song appears to have induced an experience too similar to love in participants for songs to show a large enough contrast.

Another way of developing novel stimuli that may seem to bypass interpretation lies at the basis of the lists of personality adjectives employed in CNS of self-reflection. Current studies refer to early studies in CNS of self-reflection, which refer back to a study by Norman Anderson (1968) as the source of their personality adjectives. In turn, Anderson selected his adjectives from lists compiled by Gordon Allport and Henry Odbert (1936). Allport and Odbert went through the entire Webster's Unabridged New English Dictionary of 1925 and jotted down all words that could possibly describe personality. They found 17953 terms in total. Anderson (1968) selected a subset of 555 adjectives, through criteria that he came up with and through pilot testing the familiarity of adjectives. For example, one of the criteria he used was to exclude items from Allport and Odbert's set that denote physical characteristics. 'Strong-minded' thus made it onto the list, whereas 'strong' did not. Anderson obtained likableness ratings for these, that is, he let pilot participants indicate on a scale from 0-6 how well they would like a person described by the adjective. Pilot participants were university students in the 1960s. Current studies employ personality adjectives from Anderson's lists and use his likableness ratings to counterbalance the valence of the adjectives employed in an experiment.

Again, including all possible stimuli appears to be a way in which to bypass interpretation. Yet the original set of adjectives was reduced to the lists employed currently in CNS of self-reflection and the criteria by which this was done have conceptual consequences. I think it is safe to say that hardly anybody working with the personality adjectives in CNS of self-reflection knows the history of the stimuli they work with. The crucial insight they thereby lack is an insight into the take on personality that is embedded in the stimulus set. They also lack awareness that the likableness ratings they use in counterbalancing may have changed since the 1960s and may always have been different for different sociodemographic groups than the one from which pilot participants came. (I return to both issues in the next section.) What is more, the fact that researchers in CNS of self-reflection are overwhelmingly studying reflection on one's personality in the first place is due to the presence of Anderson's stimulus set and to a landmark study such as Craik et al. (1999) employing this set. Had Allport and Odbert not compiled their lists in the 1930s, current CNS of self-reflection would likely look differently.

All in all then, researchers have several strategies for choosing the task they employ in their study. First and foremost, they try to employ a task that already exists. When a field of study is new, as CNS of love and CNS of self-reflection were not so long ago, researchers may try to use an existing task from an adjacent research literature, optionally adapting it a little. Researchers may also develop a task themselves. They can simply think something up and argue for their choice. As far as stimuli are concerned, researchers may include many or even all possibilities in an initial set and then reduce this set by means of criteria or through pilot testing. In the latter case, the items that produce results most reliably are included. Although pilot studies may give off the impression as if researchers are empirically testing the quality of their tasks, they do not rely any less on interpretation as simply thinking up a task does. Interpretation in task development is inevitable. Without critical reflection, this may have problematic consequences, particularly for the validity of tasks.

### **Validity issues with current task development**

Current task development practices in cognitive neuroscience are not focused on the validity of tasks, nor on the conceptual implications of setting up a task in one way rather than another. I see four types of problems that this leads to. We encountered them in the first two chapters and can now understand them more systematically, as validity problems.

First, tasks may be invalid. An early study in CNS of self-reflection had a face recognition condition as a task capturing 'self processing'. Is this a picture of your face

or of the face of a stranger? participants were asked (Kircher et al. 2000). One may wonder whether any self processing is necessary to complete this task at all. Could not lower-level visual processing do the job? (Oh, there's a broad nose: left button; narrow nose again: right button). Even if there is some implicit recognition by participants that they are looking at a picture of themselves, it is questionable whether the processing of pictures of one's face is truly processing of self. After all, we experience ourselves as selves much more often than, and differently from, those few, brief moments we encounter our face in the mirror each day. And at those brief mirror moments, we meet our selves almost as we encounter others: by witnessing outer appearances. Lack of validity is also something to watch out for when tasks that were developed in one context are taken into a different context.

Second, a highly prevalent risk of current task development practices is muddled label-task connections. As mentioned in chapters 1 and 2, progress in CNS is made mostly through compiling a number of comparable studies and checking what neural activity patterns consistently correlate with the process of interest. Review articles and databases are set up to this end. As also seen in chapters 1 and 2, however, these efforts are complicated considerably when researchers use different labels for the same task or when they use the same label for different tasks. In CNS of love, maternal love and maternal attachment studies are not reviewed together due to the labelling difference, even though they employ the same task. In what I call CNS of self-reflection, a wide variety of labels is used: besides 'self-reflection', one can find 'self-knowledge', 'self-awareness', 'self-referential reflective activity', 'self-referential processing', 'thinking about selves', 'metacognitive evaluation of the self', 'judgments of self', or simply 'the self'. Although the variety of labels suggests a variety of tasks employed and processes tapped into, they refer to two tasks only: one involving free reflection on one's personality; the other revolving around quickly accessing a rough and ready understanding of one's personality. Review articles in CNS of self-reflection tend to include most or all of these studies, thereby lumping together two different conceptions of self-reflection. If validity refers to tasks inducing the process or experience they purport to be inducing, muddled label-task connections affect validity across studies, as different researchers call the process they purport to be inducing with the same task by different names, or use the same name for a process they aim to be inducing with different tasks. Researchers are not specific nor systematic enough in setting up label-task connections in CNS research. This corrupts the quality of review articles and databases.

Thirdly, the validity of CNS research may suffer when biases are built into a task. In chapter 1 we saw how CNS of love has a disproportional number of female over male participants, which may in part be due to socio-cultural biases. CNS of self-reflection

employs adjectives taken from Anderson's lists (1968). I mentioned before how he excluded physical characteristics, thus embedding an understanding of personality as being not of the body (and the implicit background assumptions of a difference between mind and body, with personality being of the mind) into the set of adjectives he came up with. Furthermore, with respect to his likableness ratings, it can be questioned whether so-called blue collar workers would be as positive about 'intelligent' 'open-minded' 'earnest' 'interesting' 'broad-minded' 'well-spoken' 'educated' 'clever' 'quick-witted' and 'brilliant', which are all in the top 50 of most likable personality descriptors. Systematic limitations built into the task can thus result from pre-existing understandings of love or self-reflection or any other process of interest. Pre-existing understandings may reflect theoretical convictions of a researcher, but they may also reflect understandings present in the wider socio-cultural environment.

Finally, any task is bound to be limited in that it is bound to cover one aspect of a phenomenon only. This is different from the previous point. In this case a task's limitations are not so much based on researchers' pre-existing understanding of a phenomenon, but simply reflect the need to manipulate one factor at a time in the experimental context. Instead of being based on socio-cultural or theoretical biases, these are coincidental limitations. It is important to be aware of these limitations nonetheless, as one cannot draw conclusions about the neural correlates of love, for example, based on one type of task only. Put differently, one task can never be a valid operationalization of love in general, but only of something more specific. Building up a body of research using one particular task allows researchers to check whether clear patterns in the data show up time and again. It is also important to diversify tasks, however, if researchers want to gain a full picture of the neural activity subserving a particular type of behavior or experience. Most importantly, researchers should always be aware of the conceptual implications of the tasks they develop, to communicate results appropriately instead of overselling them.

### **Decoding and other analysis developments do not circumvent issues of validity**

CNS of love and CNS of self-reflection mainly consist of fMRI studies relying on subtraction paradigms ('neural activity during experimental task' minus 'neural activity during control task' equals 'neural activity correlating with process of interest'). Studies ask localization questions mainly. Other subfields of CNS research use different types of analysis methods than those used in CNS of love and CNS of self-reflection. Researchers investigate networks instead of regions, for example, and some perform studies using decoding paradigms. Although it may not seem so at first sight, they struggle with the same issues of validity.

First, whatever neural correlate a study investigates, be it a region, a chemical, or a dynamic network, researchers aim to find a pattern of activity that is specific to the process of interest. They hope to be able to draw inferences about the structure of the mind from the neural activity. The analyses I develop here thus also hold for CNS studies in other fields investigating neural activity in dynamic networks. If researchers in CNS of love and CNS of self-reflection would start focusing their studies on networks instead of regions, they would still have to grapple with issues of validity.

Second, a very exciting development in CNS research is the development of decoding (for a quick overview, see Poldrack 2011), which is a different way of drawing inferences about the mind on the basis of imaging data than informal reverse inference is. Decoding relies on machine learning. For example, statistical machines are fed a batch of training fMRI data and learn to associate it with labels or with visual images. When they are fed new fMRI data later, they can predict with a measure of accuracy how that data should be labelled or what visual image is being processed by the brain that was scanned. I expect the use of decoding to drastically increase over the coming years. If it does, this would only underscore the relevance of my analyses, however. Decoding can be seen as a formal type of reverse inference (Poldrack 2011). It relies entirely on the training data and the way in which training data is categorized and labelled. Issues of validity and of consistent labelling are crucial. If you would like a decoder to recognize neural activity patterns correlating with love experiences, for example, it would be unlikely to succeed if you were to feed it all the data included in the review by Stephanie Ortigue and colleagues on CNS of love (2010). After all, that review includes studies on romantic love, maternal love and unconditional love, and thus all the different neural activity patterns associated with those different experiences. Decoding practices thus underscore the need for careful task analysis and careful description of tasks.

### **Databasing cannot address issues of validity**

Progress in CNS in general depends on the validity of tasks and on appropriate labelling of tasks for the psychological processes and experiences they tap into. These are “conceptual challenges that [...] remain widely underappreciated within the neuroimaging community” says Russell Poldrack, the cognitive neuroscientist who is most prominent in putting fundamental issues on the agenda (2016). Poldrack puts his hope in informatics. Large databases of neuroimaging data such as Neurosynth (<http://www.neurosynth.org>), BrainMap (<http://www.brainmap.org>), or NeuroVault (<http://www.neurovault.org>) can be used to calculate the relative probability with which the presence of particular neural activity indicates the involvement of a particular psychological process or experience. Poldrack notes, however, that “one general

problem for virtually all extant meta-analytic databases is the lack of psychologically detailed annotations” (Poldrack and Yarkoni 2016, 598).

So how do these databases deal with the conceptual side of CNS research? BrainMap allows you to calculate the strength of inferences for a very general concept such as ‘language’, but not for anything more specific, like word production or sentence comprehension in a second language. Each data set that is entered into the database has to be annotated manually for which psychological processes or experiences are involved. In contrast, Neurosynth automatically extracts the full text from research articles. It is fairly well able to predict the likelihood of a (psychological) term occurring in the paper on the basis of regional activity and thus can also be used to calculate the probability with which a reverse inference can be made (Poldrack 2011).

Both databases depend for their quality on the appropriateness of the terms to describe the psychological processes or experiences with. The main problem Poldrack sees with this is that researchers are not consistent in their use of terminology. We saw how many different terms are used in CNS of self-reflection, for example. In other literatures, researchers often define psychological terms according to the psychological theory they believe in. Poldrack appears to hope that informatics will solve this issue too, and he is collaborating with others to develop a cognitive ontology in the Cognitive Atlas project (Poldrack et al. 2011; see also Hastings et al. (2014) for similar projects). In it, psychological terms are entered, as well as the relations between them. The Cognitive Atlas project aims to provide consensus definitions, such that researchers will align the terminology they use, enabling automatized database searches. It also aims to link the psychological concepts to particular tasks and task characteristics.

This is where my work comes in. As much as I agree with most of Poldrack’s analysis of CNS’ problems and as much as I applaud all his work at overcoming them, none of his informatics solutions can circumvent the moments of interpretation on which CNS relies. Poldrack himself notes that “it is clear that careful human consideration and annotation of neuroimaging data remain critical components of most investigations and are unlikely to be replaced by machine learning approaches soon” (Poldrack and Yarkoni 2016, 607). I think Poldrack underestimates how tricky the interpretation is that is involved in annotating imaging data in terms of psychological concepts. Poldrack thinks that ‘annotation’ (the word itself sounds so innocent) is mainly problematic as it involves a lot of work; he proposes setting up a sort of wiki-website allowing researchers to “manually validate, annotate, and tag all data presently in Neurosynth” (Poldrack and Yarkoni 2016, 607). This only adds another moment of interpretation to CNS research practice, however. What if one researcher wants to tag data with a psychological concept that stems from the theory she espouses, whereas another researcher would prefer a different tag, corresponding to a different theory? And

what if researchers link the task employed in a particular study to a psychological process that the task in fact does not quite capture, as in the case of self-reflection as a label for face recognition of own face? Socio-cultural biases embedded in tasks are not weeded out if researchers merely annotate existing tasks without critically examining them. Setting up databases involves many decisions, all of which involve interpretative choices. At the moment we have conceptually messy CNS research literatures and hope that databases will help sort them out. What we should prevent is that we set up conceptually messy CNS databases.

What is more, databasing may not only relocate CNS' problems with interpretation, it may even exacerbate them. Aligning vocabularies is a great idea, but it may have a flipside, when people forget that any term (or 'tag' or 'annotation' or 'label') reflects an interpretative choice, rather than an empirical fact. Poldrack himself points the danger out for Neurosynth: given that Neurosynth mines the terminology that is used in research papers, it may reify the informal reverse inferences and unwarranted reverse inferences that researchers have made there (Poldrack 2011). That is, if enough researchers have hypothesized that a particular neural activity pattern may be due to the presence of e.g. self-reflection, then, irrespective of whether this is an appropriate interpretation of the neural data or not, Neurosynth will predict an increased probability that that same neural activity pattern correlates with self-reflection.

There is no way around moments of interpretation in the CNS research process. Any experiment requires a task and in any task conceptual assumptions are embedded. Current methods for developing and labelling tasks do not ensure the validity of tasks, nor the comparability of studies. Databases do not solve these issues, but only reinforce the need for self-critical reflection on operationalizations and the conceptual assumptions embedded in them. As philosophers are trained extensively in reflective analytical skills, they may have something to contribute.

## **2. Conceptual review as philosophy for cognitive neuroscience**

So how can the cognitive neuroscience research community weed out invalid tasks and biases built into tasks? How can we align tasks and psychological concepts such that the connections between tasks and their labels are no longer muddled, but clear? And what is the best approach to interpreting results? How can we benefit from databases but ensure that they do not repeat the conceptual issues that lead us to build up databases in the first place? I think reflexivity is dearly needed and I think philosophy has something to offer on that front.

### **Conceptual review of tasks**

I would like to propose a novel type of review for CNS: conceptual review. A conceptual review offers an analysis of the tasks employed in an entire CNS subfield and of their conceptual implications. By carefully analyzing participant selection practices, the instructions participants are given and the stimuli employed in the tasks participants are instructed to perform, it is possible to tease out conceptual assumptions regarding the psychological concept under study. These can subsequently be analyzed and critiqued where necessary, pointing the way forward for task development in that CNS field.

Whereas ordinary review articles in CNS serve to establish to what extent neural results from different studies converge, conceptual review articles would provide a necessary preliminary step. They give an overview of the different ways in which the topic of interest is operationalized and the conceptual assumptions embedded in these operationalizations. Ordinary reviews can make use of this information to compare neural results from studies that are in fact comparable and to understand how studies that use somewhat different tasks relate to each other. Performing a conceptual review is also quite different from performing a conceptual analysis, as it is done in philosophy. Conceptual analysis is an *a priori* activity: philosophers try to establish the necessary and sufficient conditions for a concept to apply and thus, ideally, to uncover *the* way in which a concept applies. In contrast, conceptual review could be called an *a posteriori* activity: after CNS researchers have developed different tasks that are supposed to tap into a particular process, the reviewer teases out the conceptual assumptions embedded in these tasks. Insofar as several tasks have been developed, conceptual review is bound to uncover several conceptualizations, which may or may not align with (aspects of) the phenomenon of interest.

In practice, part of the work I do in chapters 1 and 2 amounts to conceptual review of CNS of love and CNS of self-reflection. Through a conceptual review of CNS of love, I find out CNS of romantic love is mostly CNS of infatuation, or, to be precise, the study of the neural correlates to the feelings one cannot help but have towards the person one is infatuated with, upon seeing a picture of their face. It becomes clear how many different psychological labels are used in CNS of self-reflection; how these (unsystematically) refer to two different tasks, one corresponding more closely to our common understanding of self-reflection, with the other tapping into a rough and ready understanding of oneself. CNS of maternal love turns out to employ tasks that are also used in CNS of maternal attachment, but these studies are not reviewed together due their different labels. Etcetera.

So how did I go about setting up these conceptual reviews and how should one proceed when performing conceptual review in CNS? Firstly, one needs to list the different choices researchers make with regards to participant selection, stimuli and task instructions for the experimental as well as the control task. Appendices A and B provide inventories of choices researchers have made when operationalizing love or self-reflection. Secondly, the conceptual implications of such choices need to be teased out. Teasing out conceptual implications can be done in its own right. It may also be facilitated by considering the history of a task, or by comparing tasks to other tasks or to external (philosophical) views on the phenomenon of interest. In turn, thirdly, these conceptual implications can be analyzed, and critiqued where necessary. Is this conceptualization of the phenomenon valid at all? In what ways is it limited? And are some of these limitations due to socio-cultural biases or theoretical preferences that researchers hold? How does this conceptualization compare to conceptualizations embedded in different operationalizations? Finally, one can draw conclusions about how the field should move forward. Certain tasks or participant selection practices or stimuli may have to be dropped. Importantly, with the help of conceptual review, much better categories and labels can be used in data aggregation efforts such as review articles and databases. Also, insight into the exact aspects of a phenomenon that current operationalizations target is a good basis for focused diversification of tasks, broadening the scope of CNS research.

The step that is probably the hardest in this procedure is the second one: teasing out the conceptual implications of the choices researchers make in setting up their experiments. As said, although it can be done in its own right, this analysis may be facilitated in several ways. First, one can examine the context in which operationalizations were developed. How did researchers develop this task and what motivated them to do so? Do the aims they had differ from the aims current researchers have that employ this task? Taking a systematic look at the history of a task helps to detect and evaluate conceptual assumptions embedded in it. Earlier in this chapter, I demonstrated this when I outlined the history of the self-reflection task and stimuli. Second, one can compare different tasks to each other. This too helps to get the conceptual assumptions into clearer view. I used this strategy when I compared the two types of self-reflection task to each other in chapter 2. A third way I see in which conceptual review of tasks may be facilitated is through confronting a task with an external conceptual view on the topic of interest. This may generate the same effect as comparing two tasks to each other, bringing conceptual assumptions embedded in a task into clearer view. For example, in the previous chapter we saw how Harry Frankfurt develops a view of love as volitional first and foremost, as consisting in practical care for the existence and well-being of the beloved. Although Frankfurt downplays the

importance of feelings too much in my opinion, the least his view shows is that there are other aspects to love besides loving feelings. His view provides a stark contrast to operationalizations in CNS of love in that respect.

Conceptual review can thus be used to overcome the risks regarding validity that current CNS runs. First, conceptual review weeds out both invalid tasks and systematic biases built into tasks, corresponding to the first and third risk that I described. Researchers in CNS of love should aim to involve more men, for example. Second, conceptual review articles bring to light how inconsistently labels are often used to refer to tasks. Reviewing thus aids the process of aligning labels and tasks: 'reflection on one's personality' may be a fine label for tasks involving free self-reflection, and a different label should be used for the other task, e.g. 'accessing an understanding of one's personality'. Importantly, these labels can also be used in databases, preventing the muddle that would result if studies involving both types of tasks would be included as 'self-reflection'. Finally, conceptual reviews present a systematic view of the ways in which tasks could be diversified. If CNS of love currently employs visual stimuli only, then researchers could diversify by including auditory stimuli, such as recordings of a beloved's voice. Instead of only investigating the peak experience of being 'truly, madly, deeply in love' they could investigate the experiences of loving romantic partners who are no longer head over heels infatuated with each other. If current CNS of romantic love treats love as a passive experience only, diversification could consist of an active condition, in which lovers have to try to self-induce feelings of love, as happened in CNS of unconditional love. And besides feelings of love, researchers may also try to study the neural correlates to love attitudes or love behavior.

Conceptual review does not exist in current CNS. Every now and again someone realizes for a particular research literature that tasks may be problematic or that task differences may be responsible for unclear data patterns (e.g. Conway et al. 2005; Richards, Plate, and Ernst 2013). This is all rather ad hoc, however. Systematic task analysis happens rarely if at all. In fact, CNS of self-reflection is amongst those CNS subfields that have received most critical conceptual scrutiny. Even there, many reviews do not address conceptual issues at all, but look into the extent to which a common pattern of neural activity emerges from studies in CNS of self-reflection (e.g. van der Meer et al. 2010; Northoff et al. 2006). Other reviews mention a conceptual issue, but none amount to a full conceptual review and, crucially, most of it does not help the field progress. Some have used the lack of neural specificity to argue for a different conception of self-experience, without systematically analyzing the conceptual assumptions regarding selfhood and self-reflection present in existing tasks (e.g. Legrand and Ruby 2009; Northoff, Qin, and Feinberg 2011). They thus change the topic rather than improve the

conversation. I know of one article that traces part of the history of two tasks used in CNS of self-reflection (Zahavi and Roepstorff 2011). It does not explicate and evaluate embedded conceptual assumptions, however. Rather, it only points to the need to do so, to argue against the idea that cognitive neuroscience by itself could study where in the brain the self resides. This is a defensive point to make and it would have been stronger, I would say, had the authors also shown the constructive contribution philosophers can make by delving up conceptual assumptions in these tasks.

Certain reviews in CNS of self-reflection do part of the work I think a conceptual review should do. One is an early review by Seth Gillihan and Martha Farah (2005). They point out how tasks up until then differed either too much between experimental and control conditions or too little. Researchers took note and these days, control conditions tend to be better. Gillihan and Farah do not question the overall validity of tasks, however, nor do they analyze the precise conceptual assumptions embedded in them. Kai Vogeley and Shaun Gallagher do not do so either, even though they too approach my idea of conceptual reviewing. They review the literature on CNS of self more generally and point out how certain studies focus on self-agency whereas others focus on self-ownership; on self in action or self in space; on the social self; or reflective aspects of self or pre-reflective aspects of self (Vogeley and Gallagher 2011). The authors aim to show that “it is an important principle for work in the neuroscience of self to say precisely which aspect or conception of self is at stake” (2011, 112) and also state that the “perplexity about the brain’s role in self-specific processes may stem from the lack of any clear correspondence between theoretical conceptions of self and how cognitive neuroscience operationalizes such conceptions” (2011, 118). That is, they point to the importance of explicating conceptual assumptions embedded in tasks, but do so themselves only in the most general of ways. What is more, they do not give pointers as to *how* one could explicate the assumptions present in tasks either.

Generally then, both people at the forefront of developing databases and those involved in review efforts point to the crucial importance of analyzing tasks. Yet hardly anyone actually performs conceptual reviewing work and, to my knowledge, nobody has outlined how to approach this type of work, as I do here. I hope my work may contribute to a change on this front as I think it should change if CNS is to progress.

### **Conceptual review facilitates front-loading phenomenology**

Conceptual review takes place after CNS researchers have performed a certain number of studies; the contribution philosophers make through conceptual review is reflexive. Philosophers may also have something to contribute before research takes place, however, or in addition to conceptual review. Others have discussed the possibility to

front-load phenomenological analyses into experimental design or to clarify concepts that neuroscientists use (e.g. Gallagher 2003; Bennett and Hacker 2003). I would like to point out that these philosophical analyses of the structural characteristics of an experience or a concept are more likely to be useful to cognitive neuroscientists if philosophers understand the current state of the CNS subfield they aim to contribute to. Conceptual review may provide such understanding.

It may seem to philosophers to be enough to develop their phenomenological and conceptual analyses. It may seem that they thereby do the necessary preliminary work to empirical studies, given that “[c]onceptual questions antecede matters of truth and falsehood” (Bennett and Hacker 2003, 2). Obviously, I agree that any empirical question involves conceptual questions; this study returns to that point again and again. However, I am not overly optimistic that leaving a phenomenological or conceptual analysis at the doorstep of cognitive neuroscientists is likely to be useful to them. First, different analyses of a particular concept exist and how is a cognitive neuroscientist to choose among them? She almost has to become a philosopher herself to be able to evaluate them on their merits. Second, and more fundamentally, it is far from easy to move from a concept to an operationalization. A concept may have way too many aspects for neuroscientists to be able to manipulate just one. Also, philosophers’ phenomenological and conceptual analyses mostly do not straightforwardly translate into an experimental *and* a control condition. Conceptual review may help to focus philosophical efforts where they are most needed. They may also help philosophers get a grasp of where cognitive neuroscientists stand and how they currently develop their experimental and control tasks. This in turn may lead to more directly useful philosophical suggestions for operationalizations.

To give just one example, CNS of self-reflection is currently focused on reflection on one’s personality. Knowing this helps to diversify operationalizations to include reflection on, for example, the existential aspect of oneself. Some of the philosophical views we encountered in the previous chapter could help to devise operationalizations. From the conceptual review of current CNS of self-reflection, we learn that we could use a paradigm similar to the one used in CNS of free self-reflection: participants are required to reflect on themselves for a minute or so. Neural activity during self-reflection is contrasted to neural activity during reflection on a close friend. Instead of requiring participants to reflect on their personality, we could give the task a more existential bend by prompting their reflections with questions drawn from, for example, Kierkegaard’s view of existential selfhood.

Participants can be asked to reflect on their existential possibilities and on their necessary limitations, for example. They could be asked to reflect on the following statements: ‘If my life were to change drastically, I could imagine being...’ and ‘Even if

my life were to change drastically, I would always still be...'. They could be asked to reflect on their self-relation and all that it entails. Questions for reflection may include items such as: 'Do you feel you have an influence over who you are? In what way?' or 'Do you feel you have had an influence over who you have become so far? Do you feel responsible for who you will become in the future?' Or with respect to attitude towards self: 'Do you like who you are? Reflect on what you like about yourself and what not (or less).' Furthermore, participants can be asked to reflect on their selves through reflecting on the fact that they have not established themselves. For example, they can be asked to reflect on parts of themselves that they would not have chosen had they had a choice, as well as their attitude towards those parts. They can also be asked to reflect on something that happened in their life that changed the way they are.

### **Conceptual review may bring to light further concepts requiring analysis**

Conceptual review may point out basic concepts used in CNS that require philosophical reflection. Currently, philosophers engaging with CNS often focus on concepts and lines of reasoning that philosophers were already thinking about. For example, they recognize the mereological fallacy as their favorite philosopher discussed it in his work. ("It comes to this: Only of a human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees, is blind; hears, is deaf; is conscious or unconscious" says Wittgenstein in his *Philosophical Investigations*, paragraph 281 (via Bennett et al. 2007, 78)). There is nothing wrong with this per se; these analyses can be very fruitful. Yet there may be other concepts and lines of reasoning present in CNS that urgently require philosophical reflection too.

Most CNS research investigates the neural activity underlying various psychological processes. Whether it is about working memory or language comprehension, these processes can be conceptualized as ways of processing information. Cognitive psychologists seek to analyze the information process into its components and figure out the way in which these components interrelate. Stereotypically, their theories consist of boxes representing component processes and arrows depicting relations between them. In total, this should give an account of the psychological processes causing the behavior someone is able to display. Cognitive neuroscientists aim to uncover the neural activity that correlates with a particular process or one of its components. When cognitive psychologists draw on CNS in the hope that it will provide further insight into the structure of the mind, they hope to learn from overlap or differences in neural activity about overlapping components between different processes. When cognitive neuroscientists need to come up with hypotheses and operationalizations, they draw these from research in cognitive psychology.

More and more, CNS researchers are studying phenomena that do not quite fit the mold of information processing. They expand the CNS research enterprise to include so-called social neuroscience and affective neuroscience. CNS of love is an example. Love cannot easily be understood as a way of processing information, like memory and language comprehension can. Participants look at a picture of their loved one. The psychological processes involved in causing face recognition do not seem to be different from those involved in looking at a picture of a friend; it is the person depicted that is different and the meaning that person has for the participant. It is far from clear that this difference can be conceived of as a process. Something similar may hold for self-reflection. Some researchers treat self-reflection as a particular type of information processing: self-referential processing. Yet other researchers describe their task as tapping into a *representation* of the self, rather than the *process* of self-reflection, or also into *self-experience*. All in all, it is not very clear what the basic category is that we are dealing with in these cases, whether it is a psychological process, or the content a process works with, or something else altogether.

This also means that it is not very clear how CNS of love or CNS of self-reflection could be connected to cognitive psychological theories. Yet without connection to theory, researchers cannot pit hypotheses against each other, falsifying or corroborating certain theories but not others. It is not clear how to move research forward in any systematic way without theory formation, yet the way in which theories generally are formed does not appear to apply in these cases. Philosophers may be able to help sort out the fundamental categories that CNS research employs. Philosophers may also help to think through whether theory formation in cognitive psychology can somehow accommodate these different categories. If it cannot, philosophers may be able to elucidate why, as well as how theory formation around these categories relates to theories in cognitive psychology. Reviewing an entire CNS field conceptually may thus point out concepts and lines of reasoning that require elucidation.

### **Improving interpretation of results**

In chapters 1 and 2 we saw how interpretation does not just play a role during task development, but also when researchers interpret their results. Researchers hypothesize about why they found particular neural activation, i.e. what type of processing that neural activity may reflect. As said before, the CNS community uses review articles and databases in the hope of gaining a clearer view on the neural activity that systematically correlates with a particular process; as well as of the specificity with which it does so. This allows one to calculate how much one's confidence in a reverse inference increases given the evidence (Poldrack 2006). Crucially, the quality of review articles and

databases depends on the quality of the tasks that are employed in the primary research and on appropriate and consistent labelling of these tasks. That is to say, without quality interpretation practices during task development and conceptual review to assist in this, it is impossible to reach quality interpretations of results.

That said, certain interpretation practices are more likely to result in quality interpretations of results than others. Currently, the incentives driving interpretation in CNS are the same as the incentives driving any type of choice researchers have to make: what do others do? Reviewers have to accept choices for a study to be published. In general, the way in which to make researchers accept one's choices is through referring to others who did the same and whose research has already been published. Adopting others' interpretations of results thus often happens. Over time, certain interpretations get standardized, at the risk of people forgetting that it is an interpretation instead of a fact and that other interpretations may be just as apt.

The main incentive to interpret results differently from other researchers is when one favors a theory that is different from the one guiding interpretation in existing studies. When two or more theories are present, researchers should preferably pit them against each other in an experiment. In CNS of love, however, researchers' divergent understandings of love do not stem from (psychological) theories so much as from their own preconceptions, such as that love is a motivation rather than an emotion, or that it is addictive. Connecting CNS of love to systematic theorizing would therefore be an improvement, as it would facilitate setting up experiments pitting theories against each other. Even then, however, data may still be interpreted in several ways and researchers may therefore still interpret along the lines of their favorite theory and not the others.

Poldrack suggests that researchers should base the informal reverse inferences they make in their discussion sections on calculations from databases rather than on their own informal reading of the literature (2011). However, given that those databases are built up of studies where researchers have made informal reverse inferences based on an informal reading of the literature and their own theoretical preferences, one risks repeating and reinforcing those limited interpretations if one relies for interpretation merely on databases. As we saw in the previous chapter on philosophy, moments of interpretation are best dealt with through (self-)critical reflection and dissent. Just as CNS would benefit from critical reflection on tasks and diversification of tasks, it would also benefit from critical reflection on and diversification of interpretations of results. It would be good if the research process were to contain incentives towards these goals, instead of incentives to standardize interpretations.

One way in which interpretation of results could be improved is through listing several possible interpretations of the data. In this way one presents the situation as it is: currently, we mostly cannot be sure what neural activity represents psychologically.

Several interpretations are possible and we cannot yet decide definitely which ones to rule out. One study in CNS of self-reflection indeed does this: Johnson and colleagues list five different possible interpretations of the neural activity pattern they found (2006). Once the possibilities are out in the open, researchers can start arguing for or against the different options, and developing experiments that pit the different interpretations against each other.

Another way in which diversification of interpretations and critical reflection on interpretations could be stimulated is through adversarial collaboration. Nobel prize winning psychologist Daniel Kahneman is the most famous advocate of this idea. Adversarial collaboration happens when researchers from opposing theoretical camps team up to investigate a particular hypothesis on which they disagree. They have to agree on a plan about how to test the hypothesis and, ideally, also on what would count as evidence for or against it. After performing the experiment, they write up a paper together. Instead of buttressing one's own theoretical positions time and again, adversarial collaboration thus force members of opposing camps to communicate constructively with one another, moving science forward. Adversarial collaboration are currently very rare in cognitive neuroscience, but I think they could be an important tool.

The most powerful impetus for research practices and changes in them comes from the criteria funding bodies use and from editorial policies of journals. If funding bodies would put a premium on adversarial collaborations, the number of them would most definitely increase. If journals would dedicate a section specifically to adversarial collaborations, they would start appearing more and more. The same holds for conceptual review articles, by the way. Furthermore, guidelines on what to report for imaging studies (Poldrack et al. 2008) could include the point that one should aim to provide as many viable interpretations of the data as possible, or at least the two strongest ones one can think of.

### **Toward a reflexive cognitive neuroscience**

Throughout this chapter, I have spoken of philosophers who could make contributions to CNS research. To be clear, in principle it does not matter who provides conceptual review of tasks, as long as it is being done, and done well. Doing this type of work well depends on having reflexive analytical skills. In philosophy training, those skills are emphasized and I am inclined to think one needs such training to be able to set up quality conceptual reviews. Shaun Gallagher is a philosopher, for example, and Martha Farah was trained as a philosopher too. I do not think it is a coincidence that they are behind the reviews in CNS of self-reflection that come closest to providing conceptual

review. Even so, possessing excellent reflexive analytical skills is not enough. One also needs a proper understanding of the CNS research process and the strict requirements that necessarily accompany experimental design, as these restrict researchers' choices.

Properly appreciating the difficulties involved in doing CNS research may also ensure that philosophers strike the right tone. In chapter 1, I quoted Raymond Tallis, who, by the way, is not a philosopher but a humanities-loving emeritus professor of geriatric medicine. Tallis dismisses CNS of romantic love by stating that "as anyone knows who has been in love – indeed anyone who is not a Martian – love is not like a response to a simple stimulus such as a picture. [...] It encompasses many things" (2011, 77). He thereby displays a lack of understanding of experimental research, which after all cannot proceed by manipulating 'many things' at once. He also manages to strike a mocking tone. The combination of these factors is likely to alienate the cognitive neuroscientists among his readers. 'Tone' may seem to be less substantial than the content of analyses, but I would venture it is of key importance in successful interdisciplinary cooperation. What is more, tone and content often go hand in hand. It is far easier to mock CNS (or philosophy, for that matter) when one does not understand it properly.

Whether it is done by philosophers who have a thorough understanding of CNS' research process or by cognitive neuroscientists who have received some training in philosophy, conceptual review and other philosophical work *for* neuroscience is key to advancement of CNS. Let us say that all the changes I propose are implemented and that CNS develops in ideal ways in years to come. What types of contributions to our understanding of the human condition may CNS provide then?

### **3. Future CNS' contributions to self-understanding**

Many arguments about the brain and the relevance of cognitive neuroscience for self-understanding contain phrases like 'in a perfect state of neuroscience...' or 'in a future neuroscience...'. In my view, the crucial question is not whether or when we will reach a perfect state of cognitive neuroscience (CNS), because such a perfect state does not exist. We won't. Cognitive neuroscience will always rely on technologies that have their limits. It will always be testing human participants and humans will always be different from one another in unexpected ways. CNS data will thus always be noisy. That is not to say that CNS will not advance much beyond its current state. It most certainly will. If the conceptual improvements I propose are implemented; if imaging technologies advance; if people come up with even smarter data analysis methods; if basic neuroscience

progresses and we start understanding more about the relation of activity in single neurons to activity in populations of neurons to activity in regions or in networks of regions; that is, if CNS evolves in all the right ways over the years to come, what would the result be? What types of insight into ourselves can we expect from cognitive neuroscience in the future?

### **Humans as embrained beings**

First and foremost, CNS is bound to advance with respect to questions about the brain-in-action. In chapters 1 and 2 we saw that current CNS of love and CNS of self-reflection address questions about the brain-in-action. What neural activity correlates with experiencing love, what neural activity correlates with self-reflection? How does this neural activity compare to neural activity correlated with closely related processes? Is there a neural correlate specific to love or self-reflection, or to one of its components? How do neural correlates to love in humans compare to neural correlates to attachment in other animals? And does correlated neural activity differ for different groups of people, for example people of different ages, or different cultural backgrounds, or different relationship statuses? We saw how the questions that studies actually address are even more specific than that, given the conceptual assumptions built into tasks. Particularly if the conceptual improvements I propose are implemented, CNS is bound to make progress on such questions.

What CNS would in effect contribute is an increased understanding of the enabling conditions of our behavior and experiences. As humans we are embrained beings, so to speak. To put it plainly, without brains, no love experiences and without brains, no self-reflection. If there were still people around who did not think that the brain mattered for our mental lives, CNS proves them wrong. In countries where medical care is widely available and the fact that we have a brain is generally known, I think it unlikely that these people exist. Anybody who has seen a loved one suffer from Alzheimer's or who has drunk rather too much alcohol at some point knows that the state of our brains matters for what we are capable of doing and experiencing. Yet that does not mean that people remember this at every single occasion. Hence my choice for the term 'embrained'. When we try to understand cognition, we may automatically think of the brain, yet need to be reminded that cognition would not work (in the ways that it does) without bodies and an environment either. 'Embodied embedded cognition' reminds us of that fact. Likewise, when trying to understand our selfhood, particularly the existential aspect of it, it may be useful to be reminded of our embrained nature. CNS may increase people's appreciation for the enabling role the physical brain plays in our lives, including the existential aspect of our lives. They may benefit from having

knowledge of the circumstances under which the brain thrives and the circumstances that undermine proper brain functioning. Oxygen, glucose and sleep suggest themselves, for example, as necessary for optimal brain performance. A greater awareness of their embained nature, may make people open a window for air, or have a proper breakfast instead of skipping it.

What is more, an increased appreciation of our embained nature may make people more aware of the automatic tendencies that ground our daily lives. We act out of habit more than we may like to think. More generally, our actions are less often preceded by conscious reflection than we may like to think. The CNS study that is probably most famous in pointing this out is an experiment by Benjamin Libet and colleagues (1983). In their experimental set-up, participants are instructed to move their finger when they feel like doing so and to remember the position of a swiftly moving hand on a clock at the moment when they feel the urge to move arise. Meanwhile, neural activity is measured with EEG. The study shows that one can detect neural preparation for the finger movement prior to the moment at which participants experienced an intention to move. This experiment is probably the most discussed CNS experiment. Libet himself appears to have worried at first that his experiment shows that humans cannot act freely. Such statements require a lot more analysis, however, both of the concept of free action and of the conceptual assumptions embedded in the details of the experiment. What the experiment unambiguously shows, however, is that neural activity is a necessary enabling condition for our capacity to act. It shows too that we do not always consciously decide what to do before we start doing it, also in those cases when we act exactly as we would like to. Insofar as people understand our capacity to act freely as a capacity that is independent of brain activity, or free actions as always involving conscious decision regarding how and when to act before actually initiating the action, CNS research may force them to think again.

We are biological creatures of habit to a larger extent than we may like to think. CNS may correct self-understanding where people's (including philosophers') self-understanding already incorporates an understanding of the brain, denying its particular relevance for human behavior and experience. It may also contribute to self-understanding where people (including philosophers) hold views on the causal factors leading up to their behavior and experiences, particularly where they deem these to involve conscious, rational decision making. Just like psychological experiments do, CNS experiments may underline the role that automaticity plays in human behavior and experience.

This may also be relevant for existential self-understanding in particular. Both loving and self-reflecting may be led by automatic tendencies to a larger extent than we are inclined to think. Philosophers such as Harry Frankfurt and Susan Wolf seem to

recognize that loving is something that we cannot help but do, towards some people at least. Frankfurt writes that “in virtue of necessities that are biologically embedded in our nature, we love our children [...] Often we go on loving them even after we have become persuaded that the love is unreasonable” (2004, 29–30).

As for self-reflection, however, Kierkegaard’s pseudonym Anti-Climacus deems people free to relate themselves to who they are and therefore entirely responsible when their self-relation is characterized by ‘despair’, to use Anti-Climacus’ terminology. He states: “because the relation [...] is the self, upon it rests the responsibility for all despair at every moment of its existence” (1980 [1849], 16). When it comes to ordinary illness, Anti-Climacus deems it “both cruel and inhuman” to say that someone brings his illness onto himself every moment that he suffers from it. Not so for despair, however: every moment people are in despair, i.e. every moment they relate to themselves in ways that Anti-Climacus deems suboptimal, they are entirely responsible themselves (1980 [1849], 16–17). (Anti-Climacus is one of the strictest, least compassionate pseudonyms. In the second part of *A Literary Review*, Kierkegaard proclaims that it matters how the environment has treated you for your capacity to relate to other human beings, particularly romantically (1978 [1846])). Being confronted with the role automatic tendencies play in our lives can serve as a corrective where Kierkegaard and his pseudonyms overemphasize human freedom and responsibility for their existential crises.

I would like to point out some limitations to this contribution to self-understanding by CNS. First, CNS is not the only science pointing out the role automaticity plays in our behavior and experience. Most of what we know about our automatic tendencies stems from psychology experiments, not cognitive neuroscience. Behavioral experiments have shown that our self-understanding may be incomplete or even at fault. For example, if we do not hold any overt racist convictions, we probably hold a self-understanding that we are not racist and probably also think that we do not act in racist ways. Yet various psychological studies have shown that people who lack overt racist beliefs often still discriminate on the basis of race when they act on auto-pilot, so to speak. On average, research participants stand further away from someone who is black compared to someone who is white; are quicker to associate black faces with negative words than positive words; and are more likely to classify an ambiguous object as a gun when they saw a black face just prior to it than when they saw a white face (“Implicit Bias & Philosophy” 2016). The psychological research thus may help us challenge and modify our self-understanding. Apparently our overt convictions and automatic behavior can be quite different. These experiments point out aspects of behavior that we display but do not notice in daily life. Where automatic tendencies and convictions come apart,

learning about the circumstances in which we are likely to act automatically versus the circumstances in which we are likely to act on our convictions may help us reduce our discriminatory behavior, if we wish to do so. We learn this from behavioral research, however, not from cognitive neuroscience research.

Second, although insight into the enabling conditions for human behavior and human experience forms a contribution to human self-understanding, there is a lot that it does not contribute. Basically, it does not contribute any further insight into the behavior or experience itself. When the enabling conditions are met, self-reflection is possible. For self-reflection to be possible, its enabling conditions should be met. However, further insight into the enabling conditions does not give us any further insight into the content of people's reflections on who they are, or into the quality of their reflections.

An example outside the sphere of existential self-understanding may help clarify this point. Say Deidre teaches mathematics to high school students to support herself while studying philosophy. One day, she explains basic functions and their derivatives. When her students ask her why she claims that the derivative of  $f(x) = 3x - 7$  is 3, she may point to the rules of differentiation. She may also explain the rationale for those rules and, for example, draw a graph and point to its slope. What she is unlikely to do, however, is to say: because my brain gets enough oxygen and glucose at the moment, and also, I have received an excellent mathematics education. If a neuroscientist were to enter the classroom and explain everything about the neural networks that get activated during mathematical problem solving, students would not gain any further understanding of functions and their derivatives. The type of understanding they are interested in is an understanding in relation to phenomena in the world and in relation to other parts of mathematics that they already understand. It is a type of understanding they will be able to use to solve mathematical problems themselves. They are not interested to hear about the underlying psychological processes and neural activity enabling mathematical problem solving. Nor, for that matter, will knowledge about Deidre's excellent teachers over the years contribute to their understanding of why the derivative of  $f(x) = 3x - 7$  is 3.

Likewise, when Anna tries to understand where her sense of dissatisfaction with her marriage stems from, she searches for insight into her relationship with her husband: what characterizes it currently, how this has changed from before, and what factors have been involved in the change. An understanding of those factors may also give her a sense of direction on how to change her situation for the better again, or at least on how to deal with her situation. Insight into the neural activity correlating with people looking at pictures of the faces of their long-term spouses is not going to help Anna gain insight into her marriage. Given that enabling conditions are met, further insight into those

enabling conditions does not provide increased understanding of the phenomena they enable.

An apparent exception is the case in which enabling conditions are not met, that is, when mathematical problem solving or loving fails consistently. If Deidre were to proclaim that the derivative of  $f(x) = 3x - 7$  is 7, she may just have been absentminded. If she fails to answer her students' questions at all or starts making gross mistakes all of the time, she and the people around her may try to understand her behavior in terms of the failure of an enabling condition. Note that enabling conditions need not be biological: if Bob were unable to explain differentiation, it may be due to an unmet enabling condition of having gone through a decent mathematics education. Given that Deidre's mathematics education has been excellent, however, something may be wrong with her brain. Likewise, if Anna consistently fails to experience pleasant emotions of any type over a longer period of time, this may also have to do with an enabling condition going unmet, including one that involves her as an embodied being. Anna may consider whether her practice of sleeping only five to six hours each night has something to do with it and start going to bed earlier. If this does not help, she may seek out some expert who can give her further insight into her anhedonia and the enabling conditions for love experiences that go unmet. Also then, however, further insight into the brain and its hormonal systems does not provide further insight into love in general or her relationship to her husband in particular, but only makes clear that an enabling condition has passed unmet.

Many philosophers have pointed out the asymmetry between, on the one hand, the lifeworld i.e. the way we engage with the world and experience the world before we start to objectify it and study it empirically, and on the other hand its enabling conditions (Husserl 1954; Sellars 1962; Strawson 1974; Zahavi 2003; Buekens 2010). Brain mechanisms play a role in both healthy and pathological behavior and experience. However, firstly, when all is relatively well and we seek to understand ourselves, we search for the reasons why we act as we do, not for the brain mechanisms that enable us to act for reasons. Anna searches for insights like 'I have changed, yet my husband's way of dealing with me has not changed accordingly and therefore I feel disconnected from him now' or 'we used to enjoy doing x, y and z together, but now we do not do any of that anymore, and therefore I feel disconnected from him'. She does not search for 'diminished activity in the midbrain dopamine network compared to happily married couples'. Secondly, deciding whether behavior and experience is relatively healthy or should be considered pathological happens at the level of the lifeworld: it is there that we decide that Deidre's explanations have become too absurd to be caused by absentmindedness alone, that some enabling condition must be unmet. Thirdly, to be able to study brain mechanisms and their involvement in enabling the behavior and

experiences that form our lifeworld, one needs to tap into that lifeworld to develop tasks, as we see throughout this study. The lifeworld is always prior to insight into its enabling conditions, even at those instances when it is relevant to consider ourselves as embrained beings.

### **Triangulation**

Insight into the brain and the brain-in-action contributes to self-understanding when it is relevant to consider ourselves as embrained beings. The hope that many people have is that CNS research may also contribute insight into the mind. Indeed they hope CNS will be able to address questions regarding “what life, mind, sex, love, thinking, feeling, moving, attending, remembering, communicating, and being are all about” (Gazzaniga, Ivry, and Mangun 2002, 1). If we assume that the conceptual improvements I propose in this chapter will be implemented, reverse inferences, whether informal ones based on databases or formal ones decoded by machines, can be made with greater reliability. Nevertheless, they will still come in the guise of probabilities, not of certainties. That is to say, the crucial issue will be to take decisions on how much weight to attach to CNS’ reverse inferences for insight into our psychologies, especially when other types of data are present.

Let me start with relatively clear-cut cases. We may choose to rely on inferences based on CNS data for insight into psychological questions when other types of data are hard or impossible to obtain. This may be the case when we study other animals, or young infants, or patients who are locked in. They may not be able to tell us about their experiences. Their ability to display behavior or to follow instructions regarding what types of behavior to display may not suffice for empirical research. CNS research may be the only avenue of investigation that is open in such cases.

We may also choose to rely on inferences based on CNS data when other types of data cannot address the psychological question at hand. For example, there are cases in which behavioral research cannot directly distinguish between two different psychological theories. If one theory predicts the same neural activity for two conditions whereas the other predicts different neural activation patterns (and certain further requirements are met (see Henson 2006)), then CNS data may increase our confidence in one theory, but not the other.

More generally, CNS research may boost our confidence in a particular psychological theory if data aligns with what we know from other types of research and what we would predict on the basis of the theory. That is, we may use CNS data for triangulation. At present, we have to be extremely careful with doing so, as my investigation in chapters 1 and 2 shows. Researchers are often inclined to interpret data

along the lines of what they already believe in. Neural data may therefore seem to triangulate nicely with a particular psychological theory, but in fact merely reflect the fact that researchers have interpreted the neural activity in line with the psychological theory that they espouse. Relying for interpretations on databases such as Neurosynth would constitute an improvement in this sense. Yet caveats remain, as databases take existing interpretations at face value and thus only come up with interpretations that are already present and prevalent in the research community. Conceptual review of CNS subfields and the employment of the findings of these reviews in databases are absolutely vital to move CNS further on these points. If they are indeed implemented on a wide scale, and continuing critical reflective attention is paid to interpretative practices, CNS data can be used for triangulation with psychological data.

What about the case in which CNS data and data from behavioral experiments do not align, but suggest different things about our psychologies? We have to take a decision then regarding how much weight to attach to either type of evidence. I would argue that we should prioritize direct evidence over indirect evidence. For the foreseeable future, it seems prudent to attach most weight to behavioral research when it comes to our psychologies and to CNS research with respect to claims about the brain, particularly the brain-during-action. It should be clear by now that interpretation is always required to let neural activity inform us about the behavior we display. Given how tentative these interpretations are we should be extremely careful in drawing those conclusions. Naturally, when behavioral research suggests one thing about our psychology and inferences about our psychology on the basis of CNS data suggest something else, we should try to devise a test pitting the hypotheses against each other. This should preferably be a behavioral test, as it could tap into psychology directly. In the meantime, as said, I suggest we put our faith in behavioral research when it comes to insight into behavior.

### **Sleepy teenagers: a case study**

When people search for self-understanding, or more generally, try to increase their understanding of human beings, this is often with an eye to learning how to deal with themselves and others better. That is to say, the search for self-understanding is not just a theoretical quest, but a practical one too, and often foremost. I would like to conclude this chapter by considering the case study of the sleepy teenagers, which was widely discussed in the Netherlands after developmental cognitive neuroscientist Eveline Crone published a popular book on the adolescent brain (2008). Popular media put it crudely: adolescents cannot help but go to bed late, because of their brains. As schools start early, they are chronically sleep-deprived. Should schools not start later?

Careful consideration of this case shows, first of all, that cognitive neuroscience has relevant contributions to make to our understanding of teenagers, precisely because it offers insight into enabling conditions that appear to go unmet during adolescence. Second, many of the crucial insights stem from behavioral research rather than research into the neural mechanisms underlying behavior. Third, research into environmental factors would provide important additional information. And fourth, normative considerations play a role at all stages of this example, as in any real-life situation in which people search for self-understanding. That is to say, this example illustrates the points I have just made and adds a final one: when searching for insight into how to deal with ourselves and others, normative considerations come into play that cannot be addressed empirically.

As always, it is vital to scrutinize the actual research for what it shows exactly. I cannot fully do so here as that would require another chapter, yet a glance at the original research drawn on in the book already reveals important information. “Our data indicate that certain aspects of the homeostatic system are unchanged from late childhood to early adulthood, while other features change in a manner that is permissive of later bedtimes in older adolescents” write Mary Carskadon, Christine Acebo and Oskar Jenni in their abstract (2004). A specific example of a feature that changes is secretion of melatonin, a hormone that facilitates falling asleep. As adolescents mature physically, the moment in the day at which melatonin is secreted becomes later. This change permits adolescents to stay up late, as the authors put it. In other words, this does not mean that adolescents *cannot* fall asleep at 9.30pm. It means that they *do not have to*, as younger children do, who simply topple over at some point or experience clear discomfort if they do not go to sleep. The authors also note that melatonin levels show large differences from individual to individual. They state that, on average, adolescents need more sleep than adults, amongst other things because of the physical changes their bodies go through. Yet schools often start early. All in all, the hormonal changes adolescents go through make it easier for them, on average, to fall asleep late. As they have to get up early to go to school, it is hard for them to get the hours of sleep they need.

CNS research thus provides insight into the physical conditions that normally enable people to fall asleep at a time that allows them to get the amount of sleep they need. This enabling condition for falling asleep at an appropriate time is compromised in adolescents. Contra some of the cruder representations of this point in the media, this does not mean it is impossible for them to fall asleep early, but it does make it harder. Whatever the causes, a study on amount of sleep in adolescents of various ages (that did not investigate neural mechanisms) shows that they indeed sleep less than they should (Carskadon, Acebo, and Jenni 2004).

Importantly, it is on the level of behavior and experience that we decide that this is a problem that needs to be addressed. Carskadon and colleagues cite US data regarding car accidents due to the driver falling asleep: just over half involve drivers aged 16-25. They also point to research showing that sleep deprivation correlates with depressed mood, poorer memory and more generally, decreased learning ability (2004). Most of these behavioral data have been obtained without taking the brain mechanisms subserving sleep into account. We could have decided there was a problem without knowing anything about melatonin. What CNS research adds is an appreciation of the biological mechanisms contributing to this problem. Now that we know they exist, we may apportion less blame to adolescents, for example, as we know it is harder for them to go to sleep early enough than it is for the rest of the population. We now also know of the biological circumstances that we need to take into account when figuring out how to address the problem.

Before we jump to the conclusion that schools should start later and/or that adolescents should be administered melatonin, there are further possibilities to be investigated. Environmental factors, broadly construed, may be able to shield adolescents against the effects of the changes in their hormonal system, just as they may contribute to the behavioral problems too. The data showing that adolescents sleep less than they should, are average data. Not all adolescents sleep too little. What is more, that most do, does not mean this cannot be avoided. After all, the changes in their hormonal system permit adolescents to fall asleep late, but do not make it absolutely necessary that they do. Good habits regarding bedtimes and not too much screen time in the hour or so before going to bed may help adolescents to fall asleep at a time that still allows them nine consecutive hours of sleep. One can think up many possible environmental factors promoting early sleep onset. If social media would have an 'adolescent mode', for example, switching off automatically after 9pm, there would be less incentive for adolescents to want to stay awake.

Again, we need to investigate the efficacy of possible measures with behavioral experiments. Do adolescents sleep more hours once a change is implemented? Do they feel less depressed and do they cause fewer car accidents? That is the information we are interested in if we want to establish whether a proposed solution for adolescents' problems indeed works. It is information about adolescents' behavior and experience. Knowing that melatonin secretion also starts earlier in the evening after a year of good habit formation would nicely triangulate with this data. If it does not, however, the relevant information is that adolescents sleep earlier and feel and function better.

There are important lessons for policy makers drawing on CNS in all of this and for neuroscientists advising policy makers too. Researchers from different disciplines are not in competition with each other, but offer different types of evidence, addressing

different questions. When drawing on CNS data for policy, it is important to complement this with behavioral data as well as with data about environmental factors affecting behavior. Philosophers may have a role to play here too, insofar as they are able to clarify how different types of empirical data relate to each other. What is more, philosophers are trained to pay critical attention to the normative dimensions that inevitably come into play in the translation from science to policy.

A final point to take away from this case then is the importance of normative considerations throughout. When researchers perform a laboratory study to investigate how much sleep adolescents need, they need to decide on a criterion for *enough* sleep. If we want to decide whether reduced sleep in adolescents is not just a non-ideal fact of life, but actually a *problem* that is serious enough to need addressing, we implicitly take a stance on when something counts as a serious problem. Do enough people suffer from this? Is their 'suffering' severe enough for it to count as a problem? And what type of evidence do we draw on to determine this? Some may be inclined to think that adolescents' depressed mood requires mitigation whereas others think depressed mood is just part of life, but falling asleep while driving is problematic. Even if adolescent sleepiness is acknowledged to be a problem, we may still think it is more important that adolescents learn to fit in with the day rhythm of society and decide not to take action. Finally, if we decide to try and tackle the issue after all, various remedies may be tested empirically for their efficacy. Yet the burdens and benefits they bring to different people need to be weighed against each other and philosophers may be of help in elucidating the arguments that can be given for and against the various remedies.

#### **4. Conclusion**

In sum, cognitive neuroscience may contribute to self-understanding in those circumstances in which it is relevant to consider ourselves as embrained beings. The type of insight it then provides is insight into the neural mechanisms that enable us to act and experience in the ways that we do. CNS data can only ever provide indirect evidence regarding behavior and experience. Under certain circumstances, that may be all we have. Under other circumstances, it may boost our confidence in a particular psychological theory. Yet it cannot trump the direct evidence that psychological, behavioral experiments can provide. What is more, when searching for an understanding of the human condition, we often wish to understand humans better so as to know how to deal with them. Normative considerations inevitably come into play then and these cannot be addressed empirically.

Cognitive neuroscience is an exciting research endeavor. The brain is still largely a terra incognita. That we are now able to discover more and more about this undiscovered territory and its involvement in human behavior and experience is no mean feat. In the introduction to this study I quoted thinkers who are skeptical about CNS' potential to contribute to human self-understanding. Neurological explanations cannot replace everyday psychological explanations of human behavior, i.e. explanations in terms of reasons, intentions and the like, they argue (Bennett and Hacker 2003). There is no way in which a complete record of all neural activity would make us find out what is 'really' going on when we act in the world or experience the world (Tallis 2011). This may be true, but it is also somewhat beside the point. Cognitive neuroscience need not be able to tell us *everything* or *replace* existing understanding for it to contribute *something* to human self-understanding, a piece of insight that is *complementary* to other types of understanding. And it does so when it points out the automatic tendencies on which much of our behavior is based. It does so where it provides insight into the conditions that have to be met for us to be able to love or to self-reflect.

Yet given that enabling conditions are met, further insight into the brain-in-action is not what we search for when we search for self-understanding. Not even in an ideal future is CNS is going to solve "big things like the meaning of life, or the meaning of meaning" as the first two editions of the cognitive neuroscience textbook had it (Gazzaniga, Ivry, and Mangun 1998, 1, 2002, 1). Maybe a dawning awareness of that fact made the authors decide to leave out those big promises in the third edition of their textbook (Gazzaniga, Ivry, and Mangun 2008).

Nevertheless, they kept the quote by Kierkegaard and their own statement opposing cognitive neuroscience to philosophy. "Ideas derived from introspection can be eloquent and fascinating, but are they true? Philosophy can add perspective, but is it right? Only scientific method can move a topic along on sure footing" (Gazzaniga, Ivry, and Mangun 2008, 4). As I noted in the previous chapter, this opposition between philosophy and CNS is unnecessary. As I hope to have shown in the first two parts of this chapter, it is also unfortunate. Given that interpretation is an inevitable part of its research process, CNS has much to gain from a closer cooperation with philosophy, closer even than many current attempts at cooperation. Conceptual review may provide the reflexive work that is needed. In turn, conceptual review facilitates attempts at front-loading phenomenology, analyzing fundamental concepts in CNS and interpreting CNS results in terms of what they suggest regarding behavior and experience.

So far I have analyzed the types of contributions we may expect from CNS and philosophy to human self-understanding. In the next chapter, I draw out the lessons about self-understanding we have encountered along the way. In effect, I outline a

philosophical understanding of existential self-understanding, synthesizing insights derived from the philosophical views in chapter 3 and improving on their flaws where necessary. Occasionally, I also draw on cognitive neuroscience and the lessons that can be gleaned from it as well as from the exposition of its limitations, in this chapter and the first two chapters. What do we search for when we search for existential self-understanding?

