Understanding existential self-understanding : philosophy meets cognitive neuroscience

Stee, M.J. van

Understanding Existential Self-Understanding

Philosophy Meets Cognitive Neuroscience

Annemarie van Stee

Understanding Existential Self-Understanding

Philosophy Meets Cognitive Neuroscience

Annemarie van Stee

ISBN 978 94 6299 620 5

Cover design: Andrea Sinke
Printed by Ridderprint BV

Understanding Existential Self-Understanding

Philosophy Meets Cognitive Neuroscience

Proefschrift
ter verkrijging van de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 21 juni, klokke 13:45 uur
door

Maria Johanna van Stee

geboren op 7 maart 1983 te Meppel

Promotor	Prof. dr. W.B. Drees
Promotiecommissie	Dr. K. Schaubroeck, Universiteit Antwerpen Prof. dr. M.M.S.K. Sie, Universiteit Leiden & Tilburg University Prof. dr. M.P.V. Slors, Radboud Universiteit Nijmegen

This work is part of the research program 'What can the humanities contribute to our practical self-understanding?', financed by the Netherlands Organization for Scientific Research (NWO), project no. 317-20-010.

Contents

Acknowledgements

First and foremost, I want to thank my supervisor, Willem B. Drees. Wim, thank you for your guidance and your confidence in me. I have learned a lot from you, not least about the way the academic world works. You set a high example in terms of your integrity, your openness and your generosity. Thank you for your availability in spite of an impressive work load and thanks also for the many, many cups of tea.

Other people have also taken the time to comment on, sometimes large, parts of this dissertation, and I have benefitted immensely from my discussions with them. Annemarie Kalis, Jan Bransen, Katrien Schaubroeck and Palmyre Oomen, thank you so much. A special thanks to cognitive neuroscientists Kirsten Weber and Sandra Langeslag for their comments.

The main context for this research has been the Horizon project, formally known as the project 'What can the humanities contribute to our practical self-understanding?', funded by the NWO. Dear Caroline, Kirsten, Odile, Pauline, Rolf and Sem, dear Annemarie, Katrien, Naomi and Wout, dear Antoine, Ingrid, Joel, Johan, Jos, Marcus, Thony and Wim, thank you for these years. I have plainly enjoyed them so much. And I have learned a lot from you, not just about self-understanding and the humanities, but also about somewhat unexpected things like whiskey, cookies and zotero. Particular thanks to Marcus Düwell as the leader of the troupe and to Jos de Mul for arranging the opportunity for me to publish an essay in *Trouw*.

One of the many benefits of working in academia is that colleagues tend to be wonderful. And I have had many colleagues along the course of this project. First of all, I want to thank colleagues at the former Leiden Institute for Religious Studies. For lunches, occasional dinners, drinks and overnight stays, I especially want to thank Irene, Markus, Merel and Tineke. Then there was the Center for Subjectivity Research in Copenhagen. Thanks are due to Arne Grøn, Dan Zahavi and René Rosfort for useful discussions. Thanks to all colleagues, but especially Adele, Alba and Yuko for making my time there so enjoyable. I want to thank colleagues at the Department of Philosophy and Religious Studies in Utrecht for their company during lunches, lectures and at the kitchenette. I especially want to mention fellow-Horizonites Annemarie, Caroline, Naomi, Sem and Wout. Thanks also to Anna and Mariske for their wonderful company during the final stretch in Nijmegen. I have presented parts of this work at conferences

and seminars in the Netherlands and abroad and am grateful for useful questions from colleagues at those occasions. Thanks to Rob in particular for ongoing discussions, whether in Antwerp, Nijmegen or Copenhagen.

Kirsten and Annemarie, thank you for being my paranymphs. Annemarie, who else could defend this thesis in case I were to faint but you? Thank you for many conversations over the years, whether on validity issues in psychology and neuroscience or on the effects of sleep deprivation. Kirsten, I may be a philosopher, but I couldn't have written this dissertation without those CNS years. What is more, I wouldn't have met you without them. Thanks for 10+ years of your lovely company.

Many family and friends deserve thanks for being in my life and distracting me from my work. They may not have anything to do with the contents of this dissertation, they do provide some of the conditions that enable me to work on it. I want to thank Sjoerd & Marije for their hospitality in Leiden. Mariske for her expert help with lay-outing. My parents, parents-in-law, Mybeth & Michelle, and Giuliana, Marjolijn et al. for regularly taking care of David. And David for being the joyful self-in-development that he is.

Last but far from least, I want to thank you, Boi, for who you are, both on your own and in relation to me, the best thing that has happened to me. Thank you.

Introduction

Anna feels vaguely yet persistently dissatisfied with her life. She has a good job, she has a husband and two children, everything appears to be just fine. And yet she wonders: is this it? She does not recognize the person that she has become. She feels like a fake, disconnected from the people around her and the things she does. Particularly her husband, if she is honest with herself. Which she is not very often; it is easier to hide her dissatisfaction, also to herself. She cannot bear herself very well during this time though. She tries to divert her attention, works long hours, takes up a hobby. Yet the dissatisfaction does not subside. One evening, she is the life of a dinner party, sharing anecdotes and getting people to laugh. Yet all the while she observes herself and wonders how long it will take before people figure out that she is just feigning happiness. Over dessert, she thinks to herself that she probably should seek out a therapist.

The most important creature in Bob's life is his dog. He shares his apartment and his days with his dog and, most days, only his dog. Twice a day, at 11am and at 6pm, Bob devotedly walks his dog. Yet just as reliably, these walks involve Bob being dragged along by the dog and tripping because of it; it involves him kicking the dog and swearing at it; after which Bob is the one who needs to drag the dog along; etcetera. Similar patterns characterize their interaction at home. In effect, Bob spends many of his waking hours being frustrated and angry. When the dog disappears, however, he is utterly distraught. He cannot make up his mind whether to go search for his dog or not. He now spends his time in his bedroom, crying.

Clemens loses his wife and three children in a car accident in France. He is the only surviving member of their family. In an interview conducted a few years after the accident, he declares: "I myself also died in that car crash. I still have my character traits, but my identity is gone. I am no longer the person that I thought I was; I have to discover myself anew."

A promising young mathematician called Deidre is becoming increasingly restless with her work. Solving complex mathematical problems does not give her the sense of fulfillment she once experienced. For her birthday she receives some books, one of them a philosophy book. Reading it, she is gripped. She searches for more and enrolls in a course. Even though most of her time is still spent on mathematics, she enjoys the

hours she spends on philosophy more. She is so excited. To her, philosophy seems like a discipline where one may apply lucid and rational reflection to fundamental concerns of human life, rather than restricting oneself to the clean, quantifiable world of mathematics. Many years later, Deidre's appreciation of mathematics and philosophy has become more nuanced. She is quite a well-known philosopher by then and recounts her professional journey to an audience of philosophy colleagues.

Edward is a butler at Darlington Hall. He is entirely devoted to his work. Dignity is of primary importance for a butler, as he sees it, and it requires serving a lord who aims to further the progress of humanity as well as having the ability to control oneself in moments of strong emotion. One evening in 1923, Edward continues to serve at a large gathering at Darlington Hall, in spite of his father lying on his deathbed. Whereas lesser butlers may have prioritized sitting with their dying father or experienced a dilemma as to what to do, Edward remained professional and therefore looks upon that evening with a sense of triumph. Reconsidering these times years later, he starts questioning this positive take on his career. Lord Darlington sympathized with the Nazis. Although he did so out of a misplaced idealism, it does reflect on the dignity of serving in his household. What is more, it becomes clear that Edward has never fully appreciated his love for Ms. Kenton, the former housekeeper. At one point, he realizes he could have spent his life with her. His heart breaks.[1]

The situations of Anna, Bob, Clemens, Deidre and Edward illustrate a particularly human phenomenon. Human beings have a sense of self that they can question and hold an understanding of. In the cases of Anna and the others, their sense of self is tied up with people and pursuits that are of great importance to them. Insofar as they aim to understand who they are, they try to figure out what is most meaningful to them, how to deal with the people and pursuits that affect them profoundly, and what may give them a sense of direction in life. Naturally, this is not the only feature of ourselves that we may gain an understanding of. Bob seems quite unaware of who he is, but he likely recognizes himself in the mirror just fine. Clemens cannot tell who he is anymore, even though he is aware that he is as generous and impatient as he has always been. Anna questions much at the moment, but not that she is the one experiencing the dissatisfaction and that it is she who married her husband some ten years ago. Selfhood in these cases is not about recognition of physical appearance or about character traits or about being a subject of experience that persists through time. To delineate that I speak of selfhood and self-understanding in a way that involves that which is of utmost

[1] These vignettes are based on the stories of real and fictional people: Anna (Smith 2012); Bob (Camus 2013); Clemens (Peters 2013); Deidre (Wolf 1997); Edward (Ishiguro 1989).

concern to people, I use the terms 'the existential aspect of selfhood' and 'existential self-understanding'.

My choice for the label 'existential' is motivated by the fact that we often refer to the questions Anna, Bob, Clemens, Deidre and Edward ask themselves as existential questions; we also refer to their crises as existential crises. In French existentialism, 'existential' is used differently: conscious deliberation and conscious choice are emphasized, often against a somewhat gloomy background view of human existence as inherently meaningless. My use of 'existential' revolves around our ties to people and projects that are of utmost importance to us. Contra French existentialism, these existential concerns also influence our actions and experiences without conscious deliberation or even awareness. Indeed, they may do so in ways that make us happy and provide meaning to our lives. The stories about the lived experiences of Anna, Bob, Clemens, Deidre and Edward should give a first impression of what 'existential selfhood' and 'existential self-understanding' and the umbrella term 'the existential aspect of the human condition' refer to. The rest of this study is devoted to investigating how to understand it more precisely.

I aim to contribute to our understanding of the existential aspect of the human condition by considering potential contributions of insight from two quarters of the academy: cognitive neuroscience (CNS) and philosophy. In the process of studying these contributions, I also investigate what types of questions these academic disciplines address; how they address these questions; and how they may or may not mutually reinforce one another.

With the advent and expansion of cognitive neuroscience, much has been made of its potential to contribute to human self-understanding. Neuroscientists themselves and philosophers who draw on neuroscience often frame the relevance of cognitive neuroscience in these terms. A major cognitive neuroscience textbook starts off as follows:

> Do you wonder about big things like the meaning of life, or the meaning of meaning? Or are you the type who does not wonder about such evanescent questions? If you are the latter, do not read this book—even though you should. This book is for those who wonder what life, mind, sex, love, thinking, feeling, moving, attending, remembering, communicating, and being are all about. And better, it is about scientific approaches to these grand issues (Gazzaniga, Ivry, and Mangun 2002, 1).

Or to take an example from a handbook on philosophy and the neurosciences:

> The Greek oracle admonished 'Know thyself.' But for more than two millennia, the only avenues to self-knowledge were to examine one's own thoughts or to review one's behavior. The idea of knowing oneself by knowing how one's brain worked was at best a philosopher's thought experiment. [...] no one imagined that by [the twentieth] century's end we would be close to realizing this fantasy (Bechtel, Mandik, and Mundale 2001, 4).

Increasingly, cognitive neuroscientists are studying topics of existential importance: love, self-reflection, free will, religious experience, moral judgment. Overview articles also argue for the relevance of this research to self-understanding. For example, a *Nature* opinion article entitled 'Love: neuroscience reveals all' contains the suggestion that perhaps "tests for the suitability of potential partners will one day become available, the results of which could accompany, and even override, our gut instincts in selecting the perfect partner" (Young 2009, 148).

In contributing to our self-understanding, cognitive neuroscience promises to base our self-understanding on a scientific footing. Many of these authors argue this understanding is therefore more secure than the understanding philosophy is able to provide. The cognitive neuroscience textbook cites the philosopher Søren Kierkegaard, who ridicules the idea that one could study 'how consciousness comes into existence' by looking through some sort of 'giant microscope'. The textbook's authors beg to differ. They write:

> Armchair thinking is a wonderful thing and has produced fascinating science such as theoretical physics and mathematics. But to understand how a biological system works, a laboratory is needed and experiments have to be performed. 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 2002, 2).

Some philosophers drawing on neuroscience voice similar ideas. "Philosophy, in its traditional guise, addresses questions where experimental science has not yet nailed down plausible explanatory theories." writes Patricia Churchland. "The dominant methodology of philosophy of mind and morals in the twentieth century was conceptual analysis. Pilloried by philosophers of science as *know-nothing philosophy*, conceptual analysis starts with what introspection reveals about the allegedly unassailable truths of folk psychology. Then, via reflection and maybe a thought experiment, you figure out what must be true about the mind." It seems only natural therefore that "[t]he history of science can be seen as a gradual process whereby speculative philosophy cedes

intellectual space to increasingly well-grounded experimental disciplines" (Churchland 2008, 409). Self-identifying as a philosopher, Churchland has turned to cognitive neuroscience for insight into the human condition. She describes the aim of neurophilosophy, a field she founded, as "*expanding* and *modifying* our self-conception through knowledge of the brain" (Churchland 2013, 20). She is far from alone in thinking cognitive neuroscience has far-reaching contributions to make to self-understanding (e.g. Farah and Heberlein 2007; Farah 2012; Monyer et al. 2004; Ramachandran 2011). Many authors appear to think that cognitive neuroscience experiments are simply more reliable than the introspections, intuitions and interpretations philosophers supposedly rely on.

Other thinkers happen to disagree with these ideas, sometimes even to the extent that they deem it unlikely that cognitive neuroscience adds anything interesting to our self-understanding. "If we would obtain a complete record of all neural activity, and we were able to see the firing state of every individual neuron, would this advance our understanding in the slightest? [...] Would we be able directly to observe human consciousness and find out what is 'really' going on when we experience the world, judge it and act upon it?" asks Raymond Tallis (2011, 84). In a section title that follows he answers his own question: "Why there can never be a brain science of consciousness" (2011, 137). A similar type of all-or-nothing-ism characterizes Maxwell Bennett and Peter Hacker's dismissal of cognitive neuroscience's contribution to self-understanding: "What it [i.e. CNS] *cannot* do is *replace* the wide range of ordinary psychological explanations of human activities in terms of reasons, intentions, purposes, goals, values, rules and conventions by neurological explanations" (Bennett and Hacker 2003, 3).

They are far from the only skeptical voices (Coltheart 2006b, 2006a; Uttal 2001). Especially in recent years, there has been somewhat of a backlash against the neuro-enthusiasm of the past decade (e.g. Burton 2013; Satel and Lilienfeld 2013). Skeptical voices in this context mostly do not react to actual neuroscientific research, but to the neurohype that surrounds it. For example, Sally Satel and Scott Lilienfeld, the authors of *Brainwashed: The Seductive Appeal of Mindless Neuroscience*, admit that their book does not focus on neuroscientific research as such, but rather "[i]t is foremost an exposé of mindless neuroscience: the oversimplification, interpretive license, and premature application of brain science in the legal, commercial, clinical, and philosophical domains" (Satel and Lilienfeld 2013, 149). Tallis and Bennett and Hacker dismiss the sweeping claim that CNS may replace other sources of self-understanding. But what about more modest claims? What about the idea that actual CNS research may contribute some interesting piece of insight into ourselves?

What is needed and what I aim to provide with this study is a dispassionate and systematic analysis of the potential contributions of CNS and philosophy to self-understanding, particularly to our understanding of the existential aspect of human lives. Three issues are at stake there. First, the questions that actual CNS research addresses and the reliability of the answers it suggests, given the research methods it employs. Little reflective attention is paid to the ways in which actual CNS studies are performed and what implications this has for the conclusions one can draw on the basis of them (cf. Francken and Slors 2014; Poldrack and Yarkoni 2016). One experiment by Benjamin Libet and colleagues (1983) on so-called 'free will' is an exception to this rule, but overall neither neuroscientists themselves nor philosophers engaging with neuroscience reflect much on the methodological choices made in CNS experiments and their conceptual consequences. In chapters 1 and 2, I review CNS of love and CNS of self-reflection on these points, as love and self-reflection are two topics crucial to existential self-understanding. I point out the relevance of conceptual review for all of CNS in chapter 4. There I also discuss CNS' potential to contribute to self-understanding.

Secondly, philosophy is at stake in these discussions, particularly the types of insight it has to offer as well as the reliability of those insights in comparison to the reliability of empirical results. Cognitive neuroscientists are gradually moving into terrain that was once thought to belong exclusively to the humanities. CNS research on topics such as love and self-reflection presents a renewed challenge to philosophers to explain the added value of their philosophical work. Any such explanation should include an exposition of the ways in which they try to safeguard the quality of their analyses. In chapter 3, I give an overview of ideas of a few philosophers interested in existential selfhood, including Harry Frankfurt, Charles Taylor and Søren Kierkegaard. At the end of that chapter, I discuss the types of questions these philosophers address and the role that interpretation plays in their research, comparing it to CNS' research process. My entire study demonstrates what philosophers can do, both when they relate to empirical sciences and aim to aid them (chapters 1, 2 and 4) as well as when they reflect on the human condition on their own terms (chapters 3 and 5).

Third, existential self-understanding is at stake in these discussions and what we search for when we search for it. The human quest for self-understanding is a perennial quest, an aspect of the human condition. I deliberately choose not to define existential self-understanding at the outset. If I would, this would steer my study towards certain types of contributions but not others. I want to keep an open mind to insights from both cognitive neuroscience and philosophy. Importantly, this is not a race with a winner and a loser. I am interested in *what* both have to offer to understanding existential self-understanding, not in who has most to offer. The phenomena I started out with are my

touchstone: the stories of Anna, Bob, Clemens, Deidre and Edward. We learn about self-understanding in the entire study; I draw out these lessons in the final chapter.

With this study, I hope to clarify debates regarding the contributions to our self-understanding that we can and cannot expect from cognitive neuroscience and from philosophy. This should not just be relevant to those involved in the debates, but also to those who are interested in the questions addressed in them. They may be cognitive neuroscientists who occasionally are invited to explain the wider relevance of their work. They may be cognitive neuroscientists who wonder whether and how philosophy could be of help in their research. They may be philosophers who are confronted with the fact that cognitive neuroscientists study topics that they are interested in too and who experience unease, or interest, or hubris with respect to it. They may be science communicators who try to balance the need for public interest with the need to accurately reflect the conclusions we can draw on the basis of CNS research. They may be policy makers who are inclined to want to base their policy recommendations on CNS research but wonder whether and how it can be done. Ultimately, gaining a clearer view on existential self-understanding and CNS' and philosophy's insights into it may be relevant to all people who try to understand themselves better and who look to academic work for understanding.

This is a philosophical study. I proceed reflexively throughout. Although I review all CNS research into love and self-reflection, I do not perform experimental work myself, nor do I review existing studies in the way a cognitive neuroscientist would. Furthermore, although it is a study in philosophy engaging with cognitive neuroscience, it does not belong in any straightforward sense to philosophy of mind. It does not put forth a view regarding the ontological connections between mind and brain. Correlations between neural activity on the one hand and behavior and experiences on the other may be reliable enough for CNS research to contribute to self-understanding, irrespective of the proper way to conceive of mind-brain relations. Indeed, most CNS research is correlational and it is taken to speak to self-understanding, as we saw above. Furthermore, my interest in existential self-understanding leads me away from typical philosophy of mind issues. My philosophical analyses are no less useful to cognitive neuroscientists for it, however, for they pertain directly to CNS experiments. In terms of elucidating what philosophical research can be, my interest in existential philosophy moves us beyond the terrain where philosophy functions as an aid to empirical science and is modelled after empirical science. I therefore like to think that my departure from what would seem to be an obvious route to take proves to be a refreshing departure. In any case, it is the necessary route for the aims I have in mind.

Research questions and outline

As said, my overarching aim is to contribute to our understanding of the existential aspect of the human condition. My approach is to look at potential contributions of insight from cognitive neuroscience and philosophy. My research question is therefore as follows:

What contributions to our understanding of the existential aspect of the human condition can cognitive neuroscience and philosophy make?

This research question involves several subquestions. First, what types of contributions do current CNS of love and current CNS of self-reflection make to our understanding of existential selfhood and existential self-understanding? Second, what types of contributions do existing philosophies of existential selfhood (by Harry Frankfurt, Charles Taylor, Søren Kierkegaard, amongst others) make to our understanding of the existential aspect of the human condition? Third, what role does interpretation play in CNS' and philosophy's research process and how do the ways in which CNS and philosophy deal with interpretation differ from each other? Fourth, what can CNS and philosophy learn from each other, whether in terms of understanding existential selfhood or in terms of how to deal with interpretative moments in their research process? Fifth, in what direction should CNS develop and what contribution to human self-understanding can we expect from it then? Sixth, in what direction should philosophy of existential selfhood develop and what does it have to tell us about the existential aspect of the human condition then?

The first three chapters examine the contributions current CNS of love (chapter 1), CNS of self-reflection (chapter 2) and philosophy of existential selfhood (chapter 3) have to offer, addressing questions 1 and 2. In the final section of chapter 3, I compare their insights and the role interpretation plays in obtaining them, addressing question 3. Self-understanding turns out not so much to be informed by current CNS research into love and self-reflection but to be informing research strategies and interpretations of data. Several philosophers provide insight into something akin to the existential aspect of the human condition, yet none seem to clarify the situations of Anna, Bob, Clemens, Deidre and Edward exactly.

Chapters 4 and 5 contain my ideas on how to move CNS and philosophy of existential selfhood forward, thus pertaining to questions 4, 5 and 6. In chapter 4, I present several ideas on how philosophers may aid cognitive neuroscience in this process. I propose that 'conceptual review' is crucial. This proposal is exemplified by the

work I do in chapters 1 and 2, but why and how it should be done in CNS more generally is discussed in chapter 4. The chapter concludes with reflections on how that improved version of CNS may contribute to self-understanding. CNS data may be triangulated with other types of data, and be relevant to self-understanding directly in those cases in which it is relevant to consider ourselves as embrained beings, so to speak. In chapter 5 I outline my philosophical view on existential self-understanding, integrating strands of thought from chapter 3, improving on them where necessary and complementing them with lessons drawn from chapter 4. In effect, chapter 5 forms a clarification of the questions that are at stake in the search for existential self-understanding. Finally, I draw some conclusions.

1

Cognitive Neuroscience of Love

Cognitive neuroscience (CNS) is a scientific discipline that investigates the neurobiological underpinnings of mental processes. It combines the methods of cognitive psychology with those of neuroscience. In doing so, CNS addresses the question how brain activity gives rise to mental processes such as perception, action, attention, and especially cognitive processes such as memory, language, and reasoning. CNS research has taken a great flight in recent decades, due to the advent of neuroimaging techniques. Before, we could only learn about the brain and its involvement in mental processes through animal studies and through post-mortem autopsies on patients whose deficits were studied while they were still alive. With the advent of neuroimaging techniques it has become possible to study healthy human brains, in action. Although most research focuses on processes such as perception, memory and language, some CNS research investigates topics that pertain more directly to existential self-understanding. CNS of (so-called) free will, for example, has received a lot of attention as potentially telling us something new about ourselves. Other CNS literatures have yet to receive such attention. In this chapter, I investigate CNS of love to determine what contributions to our understanding of existential selfhood it has to offer.

CNS research specifically targeting something called 'the existential self' or 'existential self-understanding' does not exist. Certain studies use the term 'existential' in their title, but they take their cues from crude versions of Heideggerian thought. For example, they investigate what neural activity correlates with people contemplating their own death (Klackl, Jonas, and Kronbichler 2012; Quirin et al. 2011). The term 'neuroexistentialism' has also been used, albeit not in actual CNS experiments but in the theorizing surrounding CNS. First, one researcher uses it to label the argument that certain neuroscientific findings make more sense if we think of humans as being-in-the-world rather than as subjects dependent for their mental life on representations in the brain divorced from the outer world (Iacoboni 2007), a view more generally labelled embodied embedded cognition. Second, 'neuroexistentialism' sometimes labels the view that we are our brain, a potentially alienating thought that may induce a renewed search for meaning (Churchland 2013; Flanagan 2009). We encounter some of the ideas behind

these views when I turn to the question what future CNS may contribute to human self-understanding in chapter 4. First, we look into empirical research literatures, however: CNS of love in this chapter and CNS of self-reflection in the next.

When it comes to things that are of existential importance to people, love provides key instances. Clemens' identity crisis has come about after he lost the people he loves most. The situations of Bob and Anna also involve their significant others, specifically the dog and the husband that they are supposed to love but currently stand in ambiguous relations to. Deidre may be said to doubt between two loves: her passion for mathematics having grown weary, her love of philosophy growing. Edward experiences crisis exactly when he realizes he has been in denial about his love for Ms. Kenton. Insight into love is bound to give insight into existential aspects of ourselves.

Cognitive neuroscience research into love has only been around for a short while. The study that is generally considered to have been the start of the field was reported in 2000, by Andreas Bartels and Semir Zeki. I conducted a search for CNS of love studies by performing Web of Science and PubMed searches with the keywords 'neuroimaging' and 'love'. I looked through the reference lists of the research and review articles I thus found to find further studies. For a study to be included it had to satisfy the following inclusion criteria: it had to be a neuroimaging study (including electroencephalography (EEG) studies) reporting original data; of love; performed on a healthy population; and reported before 2015. I excluded studies that had 'love' amongst their key words, but looked at the modulating effects of love on e.g. attention or memory processing, as these studies ultimately aim to contribute to our understanding of attention or memory, rather than to our understanding of love. In November 2016, I made an additional search, and found two more recent studies of a similar kind. A total number of 16 + 2 studies were found. Details of these studies can be found in appendix A.

In this chapter, I review CNS of love with an eye to the questions regarding love it addresses and the conceptual assumptions about love that are embedded in choices made in the research process. In effect, this chapter is a conceptual review of CNS of love. In the first part, I walk through the set-up of a CNS experiment, following the order in which it is reported in CNS research articles: introduction, methods, results, discussion. I pay particular attention to interpretative choices researchers have to make, which generally receive little scrutiny. In the second part of the chapter, I tease out and analyze the conceptual implications of these choices. Finally, I draw conclusions regarding the questions CNS of love addresses and the role interpretation plays in its research process.[2]

[2] Parts of this chapter are included in a paper on cognitive neuroscience and self-understanding (van Stee forthcoming).

1. CNS experiments on love and the use of interpretation

Introduction: questions CNS of love aims to address

In the introduction of CNS research articles, researchers position their study within the wider literature. They introduce the question they aim to address and argue for its relevance to the wider research enterprise. Both the question itself and the framing within the wider literature provide information regarding the questions researchers aim to address with their study. If, for example, the research question is introduced as 'what are the neural correlates to early-stage intense romantic love?' and researchers argue for the relevance of that question by reference to neuroscientific literature studying pair bonding in other animals, then we know that researchers are also interested in the further question how the neural underpinnings of human romantic love compare to those of pair bonding in other animals. At the end of the article, in the discussion section, researchers return to the wider frame and discuss the implications of the results they found for those further questions.

All studies in CNS of love ask for the neural correlates of love, be it romantic love (Acevedo et al. 2012; Aron et al. 2005; Bartels and Zeki 2000; Fisher et al. 2010; Guerra et al. 2011; Kim et al. 2009; Langeslag et al. 2007; Langeslag, van der Veen, and Röder 2014; Ortigue et al. 2007; Tiedt et al. 2014; Xu et al. 2011; Zeki and Romaya 2010), maternal love (Bartels and Zeki 2004; Noriuchi, Kikuchi, and Senoo 2008) or unconditional love (Beauregard et al. 2009). All are interested in one way or the other to determine to what extent these neural correlates are specific to love or shared with other processes. From the choice of contrast condition (about which more in the next section) and the wider theoretical frame, I can deduce researchers' interest in several sets of questions.

A first set of questions tries to delineate neural activity correlated with love from neural activity correlated with processes closely related to love, so as to determine what neural activity may specifically underlie love. Research questions include:

- How do the neural correlates for love compare to those of friendship? What are the overlap and differences between the neural correlates (Acevedo et al. 2012; Bartels and Zeki 2000, 2004; Kim et al. 2009; Langeslag et al. 2007; Langeslag, van der Veen, and Röder 2014; Ortigue et al. 2007; Tiedt et al. 2014; Zeki and Romaya 2010)?
- How do neural correlates overlap or differ between love and familiarity (Acevedo et al. 2012; Aron et al. 2005; Bartels and Zeki 2004; Fisher et al. 2010; Guerra et al. 2011; Xu et al. 2011)?
- How do neural correlates of romantic love compare to those of sexual attraction (Diamond and Dickenson 2012; Ortigue et al. 2010)?

- How do neural correlates to romantic love overlap with or differ from neural correlates to attraction to beauty? That is, does processing pictures of a romantic partner differ from processing pictures of a beautiful person of the same sex and age as the romantic partner (Langeslag et al. 2007)?
- How does neural activity correlating with romantic love compare to the neural activity correlating with emotional arousal more generally (Guerra et al. 2011; Vico et al. 2010)?

In a second set of questions, researchers try to delineate the neural correlates to different forms of love from each other. Besides the review by Stephanie Ortigue and colleagues (2010), primary research articles address questions like:

- To what extent do neural mechanisms involved in romantic pair bonding overlap with or differ from attachment mechanisms in parent-infant bonding (Acevedo et al. 2012; Bartels and Zeki 2004)?
- How do neural correlates of unconditional love differ or overlap with neural correlates correlated with parental or romantic love (Beauregard et al. 2009)?
- How do neural correlates differ or overlap for romantic love versus a passion for a pursuit (Ortigue et al. 2007)?

A further set of questions seeks to explore love in terms of motivation and/or argue for a characterization of love as a motivational state:

- What are the neural mechanisms of the motivational force of love (Aron et al. 2005; Bartels and Zeki 2004; Beauregard et al. 2009; Ortigue et al. 2007)?
- Is romantic love better characterized as an emotion or as a motivation (Aron et al. 2005; Fisher et al. 2010)?

Now that research is underway, researchers ask questions to uncover more about neural activity involved in love and its proper functional interpretation.

- Can the neural results generally found in CNS of love be replicated through the use of magnetoencephalography (MEG) (Tiedt et al. 2014)?
- Do the neural results align with peripheral measures, such as heart rate, skin conductance (i.e. sweating) and electromyography (EMG) of the zygomatic major muscle (i.e. the muscle mediating the smile) (Guerra et al. 2011; Vico et al. 2010)?
- How should enhanced late positive potentials be interpreted? Do they reflect familiarity or positive emotion or something else?

Researchers may also compare different groups of participants to see whether neural correlates to their love experiences are different or the same. This has so far only been done for romantic love:

- Do neural correlates of romantic love differ for males versus females (Tiedt et al. 2014; Zeki and Romaya 2010)?
- Do they differ for homosexuals versus heterosexuals (Zeki and Romaya 2010)?
- Do neural correlates for romantic love differ in Western (American) vs. Eastern (Chinese) people (Xu et al. 2011)?
- Do neural correlates for romantic love differ when people hold traditional cultural values as compared to modern cultural values (Xu et al. 2011)?
- What are the differences and similarities between neural correlates to early-stage romantic love and neural correlates to (reported) long-term romantic love (Acevedo et al. 2012)?
- Do neural correlates to romantic love differ for people who are in a happy relationship with the person they are in love with versus for people who have recently been rejected but are still in love (Fisher et al. 2010)?

Although no study directly investigates the question of the overlap or difference between neural correlates underlying human love and equivalents in other animals, several studies are framed in those terms:

- How do pair bonding and motivational mechanisms in humans compare to those in other monogamous animals (Acevedo et al. 2012; Aron et al. 2005; Fisher et al. 2010)?
- How do mechanisms involved in maternal love in humans compare to mechanisms involved in maternal behavior in other animals (Bartels and Zeki 2004; Noriuchi, Kikuchi, and Senoo 2008)?

One study presents the beloved in different ways and compares neural activity patterns in response to those changes:

- How does neural activity differ when mothers watch their beloved child be happy or when they watch their beloved child be distressed (Noriuchi, Kikuchi, and Senoo 2008)?

Finally, studies may inquire after (and aim to predict) changes throughout the course of a love relationship:

- Do the neural correlates correlated with romantic love change over the course of 6 months and does this change correlate with a change in the experience of love (Kim et al. 2009)?

- Does brain activation in early-stage romantic love reliably correlate with (and therefore predict) later satisfaction in the relationship (Xu et al. 2011)?

Methods: setting up experiments in CNS of love

To actually address their research questions, cognitive neuroscientists have to operationalize love. They make choices regarding what participants to recruit. They choose what tasks to make participants perform whilst their brains are being scanned, which in turn involves choosing or developing stimuli and instructions, and choosing or developing control tasks. They also make choices regarding what brain imaging technique to use and how to set its parameters, and what analysis protocol to employ on the eventual data. All these choices are reported in the methods section of research articles. In this section, I summarize the choices researchers in CNS of love make that have conceptual consequences: selection of participants; choice of stimuli; wording of instructions; and choice of control task. The questions that are directly addressed in CNS of love are more specific than the ones given in the introduction of research articles. Methodological choices ensure that studies zoom in on an aspect of love. Later, in the second part of this chapter, I draw out the precise conceptual implications of methodological choices made in CNS of love.

Participants

Researchers have to select participants for their studies, strictly speaking in such a way that their pool of participants forms a representative sample of the population at large about which they aim to discover something. For pragmatic reasons, participants in experiments tend to be undergraduate students of the university at which researchers work: they are easily accessible, tend to be healthy and can be induced to participate through receiving course credit or a small amount of money in return. What is more, their homogeneity as a group increases the likelihood of finding significant results, as noise in the data due to individual differences between participants is likely to be less than between people of different cultural and educational backgrounds. It can be questioned to what extent these WEIRD people (i.e. people from Western, Educated, Industrialized, Rich, and Democratic societies) are representative of the overall population (Henrich, Heine, and Norenzayan 2010). Generally though, researchers write up their research articles as if they are.

CNS of love is interesting for the way in which researchers select participants to their experiments. Whereas for most CNS studies any healthy undergraduate will do, researchers in CNS of romantic love demand more. 'Are you truly, madly, deeply in love?'

is what the poster asked with which Bartels and Zeki (2000) recruited their participants. Reporting 'yes' in response to this question is not enough. Researchers follow up with interviews and other checks and they are strict too. Acevedo and colleagues (2012) report excluding approximately 40% of the people who professed to being madly in love. Bartels and Zeki (2000) included only 17 out of 70 participants that signed up. For CNS of maternal love on the other hand, matters are not that strict: being a mother will do. For unconditional love, finally, a procedure was used to select experts. Specifically, directors of living communities for people with severe intellectual disabilities were asked to select those employees that they deemed most capable of unconditional love. Researchers thus make different choices as to how to determine whom to include, depending on the type of love they are investigating. They do not report reasons for choosing participants in the particular way that they do. See table 1 for an overview over the ways in which participants are recruited and the characteristics that participants have.

Table 1: Participants in Cognitive Neuroscience of Love

Study	Selection criteria	Female / male	Age range (mean)	Relationship length (mean)
Acevedo et al. (2012) Long-term romantic love	Flyer, newspaper ads, word of mouth: 'Are you still madly in love with your long-term partner?' Phone screening: Relationship length >10 years Monogamous heterosexual relation Feelings of intense romantic love	10 ♀ / 7 ♂ equals ± 60% of original respondents	39-67 years (53 years)	'being married' 10-29 years (21 years)
Aron et al. (2005) Romantic love	Flyer, word of mouth: 'currently intensely in love'	10 ♀ / 7 ♂	18-26 years (20.6 years)	'being in love' 1-17 months (7.4 months)
Bartels & Zeki (2000) Romantic love	Posters, via the internet: 'truly, deeply and madly in love' Written statement describing how much they are in love + interview	11 ♀ / 6 ♂ out of 70 original respondents, 75% of which female (write B&Z)	21-37 years (24.5 years)	(2.4 years)
Bartels & Zeki (2004) Maternal love	Posters in nurseries: being mother	20 ♀	27-49 years (34 years)	'age child' 9 months – 6 years (24.4 months)

Study	Selection criteria	Female / male	Age range (mean)	Relationship length (mean)
Beauregard et al. (2009) Unconditional love	Directors of l'Arche residential communities where people with intellectual disabilities live together with so-called assistants helped recruit those assistants: 'with a very high capacity for unconditional love'; who understand the meaning of 'unconditional love'; and find work at l'Arche very gratifying.	9 ♀ / 8 ♂	20-63 years (36 years)	
Fisher et al. (2010) Rejected romantic love	Flyer, word of mouth: 'Have you just been rejected in love but can't let go?' Interview: mixed emotions obsessively thinking about beloved rejecter	10 ♀ / 5 ♂	18-21 years (19.8 years)	4-48 months (21 months) end relationship: 1-32 weeks ago (9 weeks ago)
Guerra et al. (2011) Romantic and filial love	Undergraduate students in a romantic relationship reside close to family and partner have a positive relationship with father	35 ♀	20-29 years (21.7 years)	
Kim et al. (2009) Romantic love	Advertisements by broadcaster: couples who had fallen in love < 100 days ago. Were fine with appearing in a documentary on love being broadcast on Korean television.	5 ♀ / 5 ♂ 5 couples selected out of ± 100 couples who volunteered	18-24 years (21.1 years)	< 100 days
Langeslag et al. (2007) Romantic love	Posters at Dutch university: in love with someone of opposite sex	9 ♀ / 9 ♂	18-34 years (21.5 years)	(12.1 months) 'love duration': 2.5-36 months (12.6 months)
Langeslag, Van der Veen, Röder (2014) Romantic love	Dutch students: in love with someone of opposite sex for < 9 months	9 ♀ / 6 ♂	18-25 years (20.8 years)	1-6.5 months (3.9 months) 'love duration': 2.5-8 months (5.1 months)
Noriuchi et al. (2008) Maternal love	Mothers of infants	13 ♀	(31.1 years)	'age infant': (16.5 months)

Study	Selection criteria	Female / male	Age range (mean)	Relationship length (mean)
Ortigue et al. (2007) Romantic love	Advertisements to students: 'individuals who are currently intensively in love' Heterosexual Dating, engaged, or married Have a favorite passion in life (e.g. science, sports, art) Report thinking about passionate hobby 60% of the day	36 ♀	(20.1 years)	'being in love': 1-60 months (15.3 months)
Tiedt et al. (2014) Romantic love	in a long-term relationship (> 1 year)	14 ♀ / 14 ♂	20-35 years (25.5 years)	12-96 months (40.2 months)
Vico et al. (2010) Love	Undergraduate students in a romantic relationship reside close to partner and four other loved ones	30 ♀	20-27 years	
Xu et al. (2011) Romantic love	Flyers to Beijing student email lists: "currently in a relationship and very intensely in love"	10 ♀ / 8 ♂	19-25 years (21.6 years)	1.3-13 months (6.5 months)
Zeki & Romaya (2010) Romantic love	Advertisements: 'passionately in love' In sexual relationship with lover Equal number of males and females in in heterosexual and homosexual relationships	12 ♀ / 12 ♂	19-47 years (26.3 years)	4 months – 23 years (3.7 years)

Most studies also used selection criteria that have to do with ability to perform the experiment and participant comparability on the neural level. Examples are: normal or corrected-to-normal vision; no use of (antidepressant) medication; no history of psychiatric, neurological or substance abuse disorders; right-handedness. As these criteria are not particular to CNS of love, I have not included them in the table.

Tasks: stimuli, instructions, control tasks

To study the neural correlates to love, cognitive neuroscientists have to operationalize love. That is to say, they have to make their participants experience love or otherwise be loving whilst their brains are being scanned. To this end, researchers develop tasks: they employ stimuli and instruct their participants on what to do. Importantly, researchers also develop control tasks. Performing a task generally involves several cognitive

processes, but only one is of interest. The control task also involves the irrelevant cognitive processes, but not the process of interest. Neural activity occurring during the control task is subtracted from neural activity correlated with the experimental task. The net result is data on the neural activity related to the process of interest only. This is called the subtraction method. It depends on several assumptions, for example regarding pure insertion of mental processes. Other methods exist too, each with its own assumptions. The subtraction method is still the most basic one, commonly used in CNS, and the only one used in the studies on the CNS of love. When devising operationalizations, researchers therefore think hard about finding a proper contrast between experimental and control task. The contrast may arise through using different stimuli for the experimental and control task or by changing the task participants are instructed to perform in response to the same stimuli.

The tasks employed in CNS of love are relatively homogenous. Most studies rely on pictures for stimuli. Participants have to bring pictures of their loved one that researchers then include in the experiment. Instructions differ only slightly: 'think about (non-sexual) experiences with this person' instructs a study on romantic love (Acevedo et al. 2012). 'View the pictures, think of the viewed person, and relax' demands another (Bartels and Zeki 2000) and a third, on maternal love, simply states that participants should 'view the pictures and relax' (Bartels and Zeki 2004). Participants thus do not have to press any buttons or otherwise respond physically; researchers hope they will simply induce an experience of love in their participants by showing them pictures of loved ones. In a variation on this general approach, Noriuchi, Kikuchi, and Senoo (2008) do not use pictures, but rather video recordings in their study on maternal love, instructing their participants to simply watch the recordings.

Neural activation during exposure to the loved one is then contrasted with neural activation during exposure to others. If there were no control task, the neural activity that would be measured would include neural activity related to the experience of love, but also to face perception, amongst other things. A control task is needed that requires participants to look at pictures of the face of someone they do not love. If pictures of random strangers are chosen for the control condition, however, net neural activity may reflect processing a familiar face versus an unfamiliar face, rather than a beloved face versus an unloved face. On the other hand, if researchers ask their participants to bring pictures of a friend they know as well as their partner, their feelings for the friend may not differ enough from the feelings for their partner to lead to neural activation patterns that are different enough to reach statistical significance. The choices involved are thus many and they are complex. In practice, researchers in CNS of romantic love use a variety of contrasts, all of which involve pictures of people of the same sex and age as the romantic partner. Some further control for familiarity by using pictures of a highly

familiar acquaintance such as a colleague or classmate (Acevedo et al. 2012; Aron et al. 2005; Fisher et al. 2010; Xu et al. 2011) whereas others (also) control for friendly feelings by using pictures of friends or beloved family members (Acevedo et al. 2012; Bartels and Zeki 2000; Guerra et al. 2011; Kim et al. 2009; Langeslag et al. 2007; Tiedt et al. 2014; Zeki and Romaya 2010) and for perceived beauty, by using pictures of people who were independently rated to be very beautiful (Langeslag et al. 2007).

Mario Beauregard and colleagues (2009) take a rather different approach. They vary the instructions instead of the stimuli between experimental and control condition. Pictures are once again used as stimuli, but all pictures are of unfamiliar children and adults, always with intellectual disabilities. In the experimental condition, participants are instructed to "self-generate a feeling of unconditional love towards the depicted person." In the control condition they are instructed to passively view similar pictures. Again, neural activity during the control condition is subtracted from neural activity during the experimental condition.

Table 2: Tasks in Cognitive Neuroscience of Love

Study	Stimuli	Instructions	Control task
Acevedo et al. (2012) Long-term romantic love	color picture partner	'think about (non-sexual) experiences with person'	Same instructions. Pics of: close friend highly familiar acquaintance low familiar acquaintance (all same sex and ± same age)
Aron et al. (2005) Romantic love	picture beloved (30 sec)	'think about events with person that were pleasurable, but not sexual'	Same instructions. Pics of: familiar acquaintance (of same sex and age) countback task
Bartels & Zeki (2000) Romantic love	color picture partner (17.36 sec)	'view the pictures, think of the viewed person and relax'	Same instructions. Pics of: three friends (of same sex, similar age, and known for at least as long as partner)
Bartels & Zeki (2004) Maternal love	picture own child (15 sec)	'view the pictures and relax'	Same instructions. Pics of: other child (known to participant, same age as own child) best friend acquaintance

Study	Stimuli	Instructions	Control task
Beauregard et al. (2009) Unconditional love	pictures of unfamiliar children and adults, all with intellectual disabilities (9 sec)	'self-generate a feeling of unconditional love towards the individuals depicted'	Similar pictures. Instructions:: 'simply look at the individuals depicted'
Fisher et al. (2010) Rejected romantic love	picture of beloved rejecter that stimulates feelings of intense romantic passion (30 sec)	'think about events that occurred with this person'	Same instructions. Pics of: familiar acquaintance (same sex, similar age and familiarity) compared (rejecter – neutral) here to (beloved – neutral) in Aron et al (2005)
Guerra et al. (2011) Romantic and filial love	picture of boyfriend picture of father (4 sec)	'view each picture for the entire time it is on screen'	Pics of: unfamiliar boyfriend of other participant unfamiliar father of other participant unfamiliar baby
Kim et al. (2009) Romantic love	5 pictures lover, taken at studio, different facial expressions. (30 sec)	[not reported]	Pics of: friend (5, taken at studio) blurred human faces gray screen Entire experiment repeated after 6 months
Langeslag et al. (2007) Romantic love	picture beloved (0.25 sec)	'focus on the fixation cross, pay attention to the faces; you'll have to answer questions about them' at 4 random moments between pictures, a question appeared: 'was the person on the previous picture your beloved / friend/ a stranger?'	Pics of: friend beautiful person (blended pics from beautiful people database) (both same sex as beloved)
Langeslag, Van der Veen, Röder (2014) Romantic love	picture beloved (10% of trials) (0.250 sec)	'respond to the target stimulus by pressing a button' Accuracy is more important than speed. [= oddball task]	Pics of: friend (10% of trials) stranger (80% of trials) (same sex as beloved, friend ± as well known)

Study	Stimuli	Instructions	Control task
Noriuchi et al. (2008) Maternal love	Video clips of own infant, either smiling or in distress. No sound. (32 sec)	[none reported]	Video clips of four unfamiliar infants, also either smiling or in distress. No sound. (wearing same clothes as own infant)
Ortigue et al. (2007) Romantic love	name beloved as prime (0.026 sec)	'indicate as rapidly and as accurately as possible whether or not an English word was presented' [= lexical decision task with subliminal priming: 40 positive emotional nouns 40 pronounceable non-words 40 blank]	Control primes noun 'passion' (passionate hobby, e.g. science, sports, art) name 'friend'/acquaintance (same sex as beloved, similar age and duration of friendship)
Tiedt et al. (2014) Romantic love	picture beloved (6 sec)	'view the photographs and imagine a comfortable, non-erotic situation related to the person presented'	Pics: 2 friends for whom no romantic feelings (yet same sex as beloved and known > 1 year, often for longer than beloved)
Vico et al. (2010) Love	5 pictures of loved ones (4 sec / 0.5 sec)	'just look at the screen and the pictures appearing in it for as long as they are on' 'try not to blink'	Pics: 5 unfamiliar people (database) 5 unfamiliar people (loved by other participants) 5 famous faces 5 unfamiliar babies
Xu et al. (2011) Romantic love	picture beloved (30 sec)	'while viewing the pictures, think about pleasurable and rewarding events (that you recalled in the pre-scan interview)'	Pics of acquaintance (same sex as beloved) 'think about neutral events (that you recalled in pre-scan interview)' Countback task
Zeki & Romaya (2010) Romantic love	6-8 pictures lover (16 sec)	'subjects were allowed to scan the images freely' 'to ensure consistent attention over time, participants were required to press a key when a circular bulls-eye prompt appeared'	6-8 pictures of friends (same sex as lover) blank screen

Pictures were often standardized by researchers, to ensure comparability between pictures of beloveds and control pictures. Standardization may involve cropping the picture to include the

head only, transforming color pictures into gray scale, etcetera. Countback tasks were used in several experiments between experimental trials, to prevent spillover of feelings from one trial to the next. I have only included them in the table when neural data during countback was included as a contrast for analysis.

Post-scan questionnaires

After the actual experiment has taken place, participants are often asked for a self-report on their experiences during the experiment. Sometimes this is done in a free-style interview manner (e.g. Acevedo et al. 2012), but often participants are required to rate the extent to which they felt love on a scale (Bartels and Zeki 2000, 2004; Beauregard et al. 2009). Optionally, they also have to rate other feelings. Mostly, this is done to verify that participants actually experienced love and to exclude participants who did not, or whose feelings of love were contaminated by other emotions. For example, Bartels and Zeki excluded one mother who reported being angry at her partner at the time of the experiment, which spilled over into the feelings she experienced towards her child. Scores regarding experienced love can also be used in correlational analyses.

Results: neural correlates of love

Results follow from an application of the experimental set-up as described in the methods section. The way in which they are reported is dictated by the way in which the research questions were introduced. Interpretative choice comes into play in the discussion section, where the implications of results are addressed, but not in the results section itself. Here, I just want to give readers unfamiliar with it an impression of what types of results may be reported in CNS of love.

Generally, the results of CNS studies provide information about the brain and its activity. This information about the brain is interesting in itself to cognitive neuroscientists. It does not speak to issues of self-understanding directly, though. A study in CNS of love, for example, likely reports that neural activity in the caudate nucleus and ventral tegmental area specifically correlates with looking at pictures of beloved people, but significantly less with looking at pictures of friends, acquaintances, or unknown attractive people. Several studies also find activity in the insula and anterior cingulate cortex and deactivation in amygdala (for reviews, see: (Diamond and Dickenson 2012; Ortigue et al. 2010).

Comparing different types of love, it seems that studies of romantic love more often detect activity in the ventral tegmental area than studies of maternal love and unconditional love, whereas studies of maternal love and unconditional love find activity

in periaqueductal grey matter that studies in romantic love do not report. These results are very tentative however, particularly given the fact that both these areas are very small and it is therefore uncertain whether activity takes place exactly there instead of right next to it. All types of love appear to activate regions in the striatum (Ortigue et al. 2010).

When comparing neural correlates to love in different types of people, several studies do not detect such a difference: neural correlates to looking at pictures of a beloved so far appear the same for women and men, homosexuals and heterosexuals (Zeki and Romaya 2010), Chinese and Americans (Xu et al. 2011). It is not clear how to interpret such an absence of difference on the neural level. Null results, as absences of differences are called, cannot falsify hypotheses.

Discussion: from brain activity to ideas about love

The discussion section of research articles forms a pair with the introduction section, much like the methods and results sections also form a pair. In it, cognitive neuroscientists discuss the implications of their results against the background of the research literatures to which they aim to contribute. Researchers focus on the results that speak most clearly to their question and compare these to results found in similar studies. In this process, they almost always also interpret what the neural activity may stand for in functional terms. That is to say, they form hypotheses regarding what the neural activity may have to say about love.

Up until here I have drawn out the choices researchers make in their research process: the questions they choose to focus on; ways in which they recruit and select participants; the stimuli, instructions and control tasks they employ. I have summarized some results in CNS of love and have now given a very general description of what happens in the discussion section of research articles. Cognitive neuroscientists are bound to recognize all that I say, even though they do not generally scrutinize these choices systematically. I want to devote the rest of this section to an analysis of the ways in which reverse inferences, i.e. inferences from neural activity patterns to ideas about e.g. love, influence issues of self-understanding in CNS. Many cognitive neuroscientists have heard of reverse inferences and their problems. That does not mean they understand the issues fully: I have come across excellent neuroscientists who think it is only an issue for localization research, not for research on networks; or who think it is only a problem in new fields of study, but not well-established ones. It does not mean either, unfortunately, that researchers think of the complications with reverse inferences every time they write up a discussion section to their own research articles. It remains an "unqualified truth" that "[m]any studies have drawn invalid conclusions" based on

unwarranted reverse inferences (Farah 2014, S26). What is more, even those researchers who understand reverse inferences and their complications very well, still appear to be unaware of the full implications of them for CNS' (in)ability to put existing ideas to the test. Let me unpack this.

Reverse inferences are inferences in the opposite (reverse) direction from the usual inference that can be drawn on the basis of neuroimaging research. Studies in cognitive neuroscience generally draw inferences of the type 'if cognitive process X is engaged (as operationalized through task comparison Y), then neural correlate Z is active'. For example, if participants experience love (as operationalized through viewing pictures of their beloved compared to viewing pictures of an acquaintance), then the ventral tegmental area is active. Reverse inferences are drawn in the opposite direction, from neural activity to engagement of a particular cognitive process: 'given that neural correlate Z is active in task comparison Y and that certain previous studies have found Z to be active when cognitive process X was putatively engaged, the activity of Z in the current study indicates that task comparison Y also engages cognitive process X' (Poldrack 2006). For example, given that the ventral tegmental area is active in the comparison between looking at a beloved versus a stranger, and that previous studies have found the ventral tegmental area to be active during reward processing, researchers conclude that the ventral tegmental activity in their study indicates that looking at a beloved versus an acquaintance also involves reward processing.

In this way, reverse inferences give the impression that cognitive neuroscience research speaks to our understanding of human behavior and human experience; in this case we appear to find evidence that experiencing love is rewarding. Strictly speaking, however, reverse inferences are invalid. They are an example of the logical fallacy of affirming the consequent. Many other cognitive processes besides reward processing may lead to activity in the ventral tegmental area. From the finding that reward processing (if p) leads to activity in the ventral tegmental area (then q), it therefore does not follow that when ventral tegmental activity is found (if q) this has to be due to reward processing (*then p). Reverse inferences could be valid if it were the case that only reward processing (if and only if p) leads to activity in the ventral tegmental area (then q). Bar full specificity, activity in ventral tegmental area increases the probability of reward processing going on, though the selectivity of neural activity for a mental process determines whether this probability increase is tiny or substantial. Determining selectivity is hard, however, and large databases are compiled to this end (about which more in chapter 4). Currently, reverse inferences cannot be drawn with any substantial confidence yet (Poldrack 2006).

Cognitive neuroscientists are often careful not to phrase the main conclusions of their research articles in the form of a reverse inference. That would constitute an

unwarranted reverse inference. Instead, main conclusions are about the brain and not about behavior or experience: about finding out that activity in the ventral tegmental area and striatum correlate with experiencing love, for example. Discussion sections of research articles generally include informal reverse inferences, however (Poldrack 2006). These inferences are made in the same way, but not treated as conclusions, but as hypotheses. Researchers aim to account for all neural activity they found in functional terms, with references to other studies that found similar activity and interpreted it in these functional terms too. Academic journals in which cognitive neuroscientists publish require them to do so. This is not strange: researchers hypothesize what neural activity they expect given the task participants are required to perform. When results come in, they account for the neural activity that was actually found, both the activity that aligns with their hypothesis and the activity they had not expected beforehand. These reverse inferences are informal in that they are intended as hypotheses to be tested further.

Nonetheless, informal reverse inferences are tricky as they can easily morph into unwarranted reverse inferences. As researchers try to back up their functional interpretations of neural activity by referring to others who have interpreted the activity along similar lines, a particular interpretation that is strictly speaking informal can become so prevalent over time within the research community that it seems like fact. For example, although the presence of a dopamine-regulated network involving many midbrain areas is rather well-established, the functional interpretation of this network as 'reward network' is a matter of convention. Based on reviews of results (Berridge and Kringelbach 2008; Lammel, Lim, and Malenka 2014), I would say 'pleasure network', 'motivation network' or also 'aversion network' may be equally (in)apt functional labels. What is more, discussion sections of research articles are not only read by cognitive neuroscientists, but also by academics from other disciplines interested in neuroscience, as well as by science journalists. Often, these readers do not realize the hypothetical nature of informal reverse inferences and understand them to be research findings. Unfortunately, cognitive neuroscientists themselves do not always properly appreciate the tentative nature of informal reverse inferences either and their research communication fails to flag the informality. Particularly when communicating with audiences who may not be as interested in the ventral tegmental area as cognitive neuroscientists themselves are (e.g. the general public, or funding bodies), unwarranted reverse inferences occur frequently (Farah 2014).

Then there is the question of the basis on which informal reverse inferences are drawn in a new field of study. To take an example from CNS of love: Bartels and Zeki (2004) found reduced activity in a wide variety of regions. They divided these regions into two sets: on the one hand, areas involved in cognition and (negative) emotions

("middle prefrontal, inferior parietal and middle temporal cortices mainly in the right hemisphere, as well as the posterior cingulate cortex") and on the other hand, areas "associated to negative emotions and to social, moral and 'theory of mind' tasks" ("amygdala, temporal poles, parietotemporal junction and mesial prefrontal cortex"). They then explain the deactivations in these regions as "suppression in the neural machineries associated with the critical social assessment of other people and with negative emotions" and suggest that "these findings therefore bring us closer to explaining in neurological terms why 'love makes blind'."

This example illustrates how researchers can draw on common sense / stereotypical ideas about the phenomenon they are studying ('love makes blind') for ideas regarding how to interpret unexpected neural activity. They can also draw on psychological theories and steer interpretation towards the theory they prefer. The Aron lab, for example, uses its studies to argue that love should be conceived of as a motivational state, rather than as an emotion (Aron et al. 2005; Fisher et al. 2010). Intricately, ideas about the type of contribution to self-understanding to be expected from CNS can also guide researchers in drawing informal reverse inferences, e.g. following the schema of natural science debunking lofty images of humanity. Researchers' personal temperament may sometimes play a role too. Helen Fisher considers love to be closely related to addiction (Fisher et al. 2010) and spends a section of her discussion focusing on CNS studies of addictions to various drugs that found neural activity in regions where her study also found neural activity. Although it may seem to an outsider (or also to a cognitive neuroscientist who is not paying enough critical attention) that these studies provide neural evidence for love making blind, for loving being a motivational state, for love being addictive, it is in fact the other way around: researchers' ideas about love steer their interpretations of the data.

A certain amount of freedom in interpreting results functionally, i.e. in drawing informal reverse inferences, is thus present in cognitive neuroscience. To some extent, researchers can choose what activity to focus on, choose what studies to compare it with and thus how to frame and label their results functionally. Reviewers have to accept these choices for a study to get published. An overall criterion is whether researchers can refer to other studies in which the same choices were made. Even though others may come up with different theories about love, after which experiments may be designed pitting the theories against each other, the practice of having to substantiate interpretations with references to others means that there is a strong drive towards standardization of interpretations. Especially where CNS research is detached from psychological theorizing, as is the case with much of CNS of love, interpretations made in early studies become corner stones that subsequent studies repeat. The risk is that it may seem as if a body of research is compiling supporting a particular idea about love,

whereas in fact what is compiling is a body of data interpreted along the same lines by researchers, even though other interpretations could be just as valid. It may seem less and less necessary to put this idea about love (or any other process) to the test. Via this mechanism, cognitive neuroscience may unwarrantedly fortify whatever understanding researchers already held, rather than adding to our understanding.

Such pernicious effects can occur in any subfield of CNS. Socio-cultural biases come into play most strongly for topics about which researchers hold convictions as human beings too. Love is a good example of a topic that features prominently in people's self-understanding. Pernicious effects are magnified when research is disseminated to the wider audience at an early stage. Again, this is most likely to occur for topics that feature prominently in people's self-understanding. It is easy to imagine a science journalist who read Bartels and Zeki's study think up a headline running 'A brain in love makes blind' with an article below it stating that cognitive neuroscientists have now found the brain mechanisms by which love makes people blind. Thus, the research that seems most relevant to our self-understanding is ironically also most prone to be biased when disseminated to the wider public.

2. Conceptual implications: a passive peak experience

Interpretation and choice play a role at various stages in CNS' research process. The previous section has shown us that they come into play in *selecting participants*; in developing *stimuli* to be used in the tasks participants have to perform; in wording *instructions* to describe those tasks; in developing *control tasks* to contrast activity during the experimental task with; and in the discussion of results, through *labelling* of results, choosing a frame of reference within which the results' implications are discussed and through both of these, by choosing what *informal reverse inference* to make about the functional implications of the results. That interpretation and choice play a role in the research process is inevitable and not a problem in itself. Each moment of interpretation has implications for the actual concept that is being studied, however. It is crucial to critically examine these conceptual implications, both to ensure that conclusions drawn on the basis of research are justified, as to be able to assess the reach of a field of CNS research and develop ideas about how to extend this reach further. Such systematic review of operationalizations and their conceptual implications happens rarely within the CNS community. As of yet, it has not happened at all for CNS of love. Let us look at some of the conceptual implications of the interpretative choices that researchers in CNS of love make.

Selecting participants: love as peak experience, love as easy, love as difficult

In CNS of love, selection of participants is a moment in the research process with important implications for the concept of love under study. Studies focusing on different types of love recruit their participants in starkly different ways. Cognitive neuroscientists interested in *romantic love* solicit participants with flyers asking 'are you truly, madly, deeply in love?'. In several studies, those potential participants are included who report the most intense feelings of being in love. The conceptual consequence of selecting participants in this way is that romantic love is treated as *a peak experience* in these studies. Even though happily married couples may also describe their experience as '(romantic) love', as we can imagine Clemens to have done before the accident, they are unlikely to be allowed to participate in this research. Research on romantic love performed so far thus becomes narrowed down to research on infatuation. It is research into the neural correlates of being truly, madly, deeply in love (and limited to those people who are inclined to embrace such descriptions of their feelings).

Selection processes in CNS of *maternal love*, on the other hand, are very different. Potential participants are required to be mothers and that is it. This only makes sense under the assumption that maternal love is *automatic and inescapable*. Participants are merely required to be female and to have a young child; they are treated as if they love their child fully and consistently, i.e. all of the time. In an interview after the experiment, however, one mother reported not experiencing love towards her child during the experiment, and her data therefore was excluded from analysis (Bartels and Zeki 2004). The existence of post-partum depression and the prevalence of ambiguities in human loving more generally both call into question the assumption that all mothers cannot help but experience love towards their child, all of the time. Implicit assumptions about maternal love embedded in current participant selection procedures may stem from, or at least are in line with, socio-cultural stereotypes about pink clouds accompanying early motherhood.

The selection procedure for participants to the study on the neural correlates of *unconditional love* is intricate. Participants are recruited amongst caretakers of mentally impaired people and only those who are most capable of unconditional love according to directors of these caretaking communities, are included. Embedded in the selection procedure, therefore, is the assumption that unconditional love is *difficult* and therefore something which *requires a special aptitude and/or (professional) expertise*. This contrasts starkly with conceptual assumptions regarding maternal love, even though maternal love is also often thought of as unconditional in the sense that it does not depend on receiving anything in return from the beloved.

A further pattern in participant selection in CNS of love is that *love* is largely operationalized as *female*. In the sixteen studies reviewed here, 243 females participated compared to 84 males. An earlier review by Ortigue and colleagues (2010) lists six studies with a total of 99 female and 21 male participants. Part of this discrepancy is inexplicable. Why did some researchers recruit female participants only (Guerra et al. 2011; Ortigue et al. 2010; Vico et al. 2010)? And why did reviewers accept this? Officially, researchers should strive for an equal number of male and female participants. The informal rule of thumb in CNS is that you aim for a group of 20 participants, with a 7 vs. 13 gender distribution as the limit of acceptability. When researchers recruit their participants among psychology undergraduates, they often find it easier to find female participants, as psychology students are more often female than male. Yet in these studies, researchers do not even seem to have tried to recruit males. What is more, the discrepancy in CNS of love as a whole is remarkable, especially considering that several studies recruited their participants outside of the university.

It seems some self-selection on the side of participants has taken place. Recruitment ads and flyers asking: 'are you madly in love?' apparently make more women than men respond. Socio-cultural biases may be at play. Generally, it is more socially acceptable for females than for males to embrace a statement like 'I'm truly, madly, deeply in love' and its associations with being swept away by one's feelings, being dependent on the beloved for one's happiness, etcetera. Indeed, the study with the largest gender discrepancy was the one in which participants had to identify with having recently been rejected in love (dumped, basically) and being unable to move on; all of which flies in the face of the self-reliance that is to a larger extent expected of men than of women. Researchers do not appear to reflect on these issues. Wording recruitment flyers in ways that are less emotionally dramatic than 'are you truly, madly, deeply in love?' may help reduce the gender discrepancy for participants in CNS of love.

Finally, part of the gender discrepancy is due to a research interest in maternal love, but not paternal love. A theoretical reason for focusing on women appears to be that researchers want to compare their research on human parental love to hormone research performed on parental behavior in other species, such as rats (Bartels and Zeki 2004). There, only females have been studied and thus a comparison to human females makes most sense. Nevertheless, human male parents are clearly able to display parenting behavior and they too experience love for their offspring. Researchers have yet to include them as participants to their experiments. Socio-cultural stereotypes regarding maternal love being special in ways that paternal love is supposedly not may be partially responsible. What is more, these stereotypes may be reinforced by CNS research. Experiments may result in neural activity patterns that are specific to mothers looking at their own child versus an unknown child. Researchers may discuss this result

as potentially being the neural underpinning of what is so special about maternal love. Yet, they have not put this idea to the test. Including males in CNS of parental love is the only way to figure out to what extent neural mechanisms subserving parental love overlap or differ for males versus females. Current research runs the risk of unwarrantedly reinforce existing ideas about mother love instead of adding to our understanding.

Tasks: love as a passive experience

CNS of love so far employs visual stimuli only, mostly pictures of the face of the beloved. In experiments, love is thus treated as an experience in *response to seeing* (a visual representation of) *the beloved*. Naturally, this is only one way in which we encounter our beloveds in real life. It makes pragmatic sense that CNS researchers start with visual stimuli: it is easier to show pictures to participants lying in a scanner than to present them with the smell of their beloved, for example. Further research could extend the range of stimuli, such as through confronting participants with the voice of their beloved.

In all studies in CNS of love except one, participants are instructed to relax and look at a visual representation of their beloved. Regardless of what cognitive neuroscientists themselves think of love, their choice for this operationalization has conceptual consequences. In it, the idea is embedded that love is *an experience, undergone passively*. Love is treated as an emotional experience outside of people's control: they cannot help whether they experience love or not. Let me unpack this.

First, love is assumed to be an experience, and more specifically an emotional experience, a feeling. Participants are not required to engage in any behavior towards their loved one, nor are their attitudes toward their loved one part of the operationalization. Second, love is operationalized as a passive experience: it happens to people or does not, in a way that is beyond their control. It only makes sense to instruct participants to just relax and look at pictures of their loved ones under the assumption that they will be unable to help that they experience love when confronted with the pictures and the accompanying thoughts of their romantic partner. Participants are also supposed to not be able to control what they feel in response to the pictures of others. This should not be love.

The operationalization by Beauregard and colleagues (2009) is interesting because it differs in this respect. They ask participants to self-generate feelings of unconditional love in response to pictures of unfamiliar, (visibly) mentally handicapped people and contrast this to a condition of passively viewing similar pictures. Clearly, if love were nothing but a feeling that happens to you, passive viewing should also induce

unconditional love, without someone being able to stop the experience. In fact, all participants report feeling love in both conditions, but they report more love in the self-generate condition. This tells us that, at least for these people (selected for their capacity to love unconditionally) in these circumstances (much of the time actively trying to self-generate feelings of love), they cannot help but love. At the same time however, the significant difference between the two conditions suggests that love can also be an activity rather than a passivity and that to some extent it can be under someone's own control.

The point here is not that current operationalizations are wrong or that cognitive neuroscientists are simplistic to think of love in these terms. I do not think for a second that cognitive neuroscientists believe they have captured all there is to say about love with their operationalization. Unlike Raymond Tallis, by the way, who looks at one study in CNS of romantic love and then critiques and dismisses a strawman version of it on the grounds 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" (Tallis 2011, 77). Indeed it does, but experimental research cannot proceed by manipulating 'many things' at once. This operationalization captures an aspect of love and researchers' choice for it is probably guided by pragmatic concerns as much as conceptual ones. The point is that the limitations of this operationalization also limit the conclusions we can draw on the basis of the research performed with the help of it, and that cognitive neuroscientists and readers of cognitive neuroscience need to be fully aware of that. They are not conclusions about love in general. Given participant selection and operationalization practices in CNS of romantic love, for example, what is investigated in these studies is the neural activity that correlates with the love one cannot help but feel towards the person one is infatuated with, upon seeing a picture of their face.

It is inevitable that operationalizing a mental process means zooming in on an aspect of that process. This happens necessarily, throughout all of cognitive neuroscience, whether researchers investigate love or moral judgment or language learning or visual perception. It is not a problem in itself, *as long as* researchers and everybody reading the research remain aware of the limits to the conclusions that can be drawn on the basis of research. Yet what happens is that researchers write 'romantic love' instead of something more precise such as 'infatuation' or 'the passive experience of infatuation' in their titles, abstracts, introduction and discussion sections. To others reading their research it may thus seem as if their conclusions pertain to romantic love generally.

Second, this has crucial consequences for setting up informative review articles and databases, both of which are powerful resources in CNS. In review articles and

databases many studies are taken together. By pooling together all data from these studies, overall neural activity patterns can be drawn out that exist in spite of different operationalization details. Accurate labelling is important to ensure that similar studies are grouped together and different studies are kept apart in reviewing efforts.

Yet what happens is that in CNS of love, for example, 'maternal love' studies are included in a review article (Ortigue et al. 2010). Other studies exist that employ highly similar operationalizations, i.e. mothers looking at pictures of their infants, but they do not include 'love' amongst their keywords or in their title (Leibenluft et al. 2004; Nitschke et al. 2004). These studies thus investigate the same process, but label it differently, which means that they are not included in review articles and not pooled together in databases. Vice versa, if databases group together studies on the neural correlates of 'love', this includes studies on romantic love, maternal love and unconditional love, just like the review by Ortigue and colleagues (2010) does, even though these types of love may be markedly different from each other, involving different psychological processes and experiences. It is vital to be aware of these differences, for they may be accountable for differences in neural activation patterns.

Third, it is important to be aware of the conceptual restrictions resulting from operationalization to keep in mind that not all of 'romantic love' is already being investigated by cognitive neuroscientists, and to develop ideas of how research could be extended. Researchers could recruit participants who report 'romantic love' but not 'infatuation', for example. Or they could ask of their participants to control their feelings of love to some extent, by suppressing them under one condition and giving them free reign under another. Studies could also not focus on the experience of love so much as on love behavior or love attitudes. Yet what happens in current CNS of love (as in other subfields of CNS) is that researchers repeat an early operationalization. The first study in CNS of love, by Bartels and Zeki (2000) aimed to address love as a positive emotion and embedded its research question and findings in scientific literature on the neuroscience of emotions more generally. They devised this particular operationalization with their aims in mind. All the others that came after them just copied the operationalization, even though some are primarily interested in the motivational aspect to love (e.g. Aron et al. 2005).

All in all then, interpretation and choice play a role in CNS' research process. This is inevitable and not a problem in itself. It is essential to be reflexively aware of the choices and their conceptual consequences, however. Currently, cognitive neuroscientists by and large are not. I return to these issues in chapter 4, when I present my ideas on how philosophy may help CNS progress.

3. Conclusions CNS of love

So, what types of understanding does current CNS of love contribute? What is its research process like and what role does interpretation play in it? And what consequences does this have for its contributions to our understanding of the existential aspect of the human condition?

The first main finding from this chapter is that interpretation plays an important role in the research process of cognitive neuroscience. Conceptual assumptions about love seep in through participant selection processes, operationalization of the research topic into a task with stimuli and instructions and contrasted to a control task, but also during discussion of results through labeling of results and choices made regarding what informal reverse inferences to draw. In CNS of love, interpretative choices have the consequence that love is generally conceived of as a passive, emotional experience in response to seeing a beloved. Operationalizations of different types of love embody further conceptual assumptions: romantic love is operationalized as the peak experience of infatuation; maternal love is treated as easy and automatic; unconditional love is treated as hard and requiring expertise.

To repeat, interpretation is an inevitable part of the research process and not a problem in itself. However, when people claim that the scientific method is the only way "to move a topic along on sure footing" (Gazzaniga, Ivry, and Mangun 2002, 2) or when they proclaim an opposition between "speculative philosophy" versus "well-grounded experimental disciplines" (Churchland 2008, 409), their claims have to be toned down somewhat. Any experiment's 'footing' or 'ground' is only as good as the conceptual assumptions on which it rests. What is crucial is critical awareness of interpretative choices and their conceptual implications and appropriate communication of research results in light of them. As said, current lack of critical awareness imperils the validity of operationalizations, the validity of conclusions drawn based on experiments, and the comparability of studies to each other. With the conceptual review I have developed in this chapter, I aim to further CNS of love on these points.

Second, the main contribution to our understanding of existential selfhood that we may expect from CNS of love is insight into the neural correlates of loving. Current CNS research addresses questions about the ways in which neural activity correlates with experiencing love. Studies first and foremost aim to delineate neural activity that correlates specifically with (a type or component of) love, be it a region, a network, a hormone or any other type of neural correlate. The verdict is still out on those questions. At the same time it has to be noted that we know a lot more than we knew a mere twenty years ago. In the future, CNS of love may contribute to existential self-understanding in

those circumstances in which we are interested in our brains and in ourselves as 'embrained' beings (more about this in chapter 4).

A subset of questions in CNS of love revolves around comparisons between two groups of participants, to see whether they show different neural activation patterns in response to seeing their beloved. It is unclear how we should interpret the findings that no clear neural activation differences are found between, e.g., homosexuals and heterosexuals, or between people who have recently fallen in love and have been in love for a long time, or between Chinese and Americans. The absence of evidence that they are different in terms of their neural activity patterns does not constitute evidence that they are the same in terms of neural activity patterns (hypotheses cannot be falsified on the basis of null results), let alone that it would constitute evidence regarding their experiences being the same (Henson 2006).

Anna, Bob, Clemens, Deidre and Edward are not pondering their brains, nor themselves as embrained beings. They are reflecting on who they are in terms of what they love. Current CNS does not have insight to offer into love directly. For that, we would have to be able to draw conclusions about behavior and experience from brain data, that is, we would have to be able to draw inferences in the reverse direction from ordinary ones in CNS. This requires knowledge about what neural activity correlates specifically with (components of) love and only (those components of) love. Current CNS of love has not found such specific neural correlates; in fact, such a result seems far off.

Occasionally, I meet people who are incredulous that CNS of love is even possible. Surely, they say, a natural science cannot investigate something as personal and subjective and meaningful as love? Well, it can, CNS of love is on its way. One has to see it for what it is, however. Experiments in CNS of love are set up to determine what neural activity correlates with (aspects of) loving. That is, researchers in CNS of love mainly aim to determine something about the brain-in-action. Experiments are not intended to replace people's personal stories about their love relationships, nor do researchers aim to replace philosophical analyses with their particular experiments. (Researchers often hold personal stories about their own love relationships and they are mostly blissfully unaware of philosophical analyses.) What they first and foremost aim to do is to complement existing insight with insight into the neural activity enabling loving (more about this in chapter 4). Dismissals of CNS of love, whether by philosophers or by members of the general public, are often based on wrong ideas about what it is. In that respect, they are not so different from inflated statements regarding CNS' contributions to human self-understanding.

This brings us to the third main finding from this chapter. To many, it may seem as if CNS contributes more to our understanding of the human condition than just

insight into the neural underpinnings of behavior and experience. Several factors play a role in producing this misleading impression. The informal reverse inferences researchers draw when they discuss their findings are often taken to be research findings themselves, particularly by readers unfamiliar with CNS' reporting conventions, but sometimes also by uncritical cognitive neuroscientists. Especially when speaking to audiences not composed of people who think the ventral tegmental area is inherently interesting, cognitive neuroscientists may cross the line and unwarrantedly present hypothetical reverse inferences as research results.

What is more, this mechanism is exacerbated by incentives in reporting and reviewing practices towards repeating what others have done, both during operationalization and for data interpretation. A particular interpretation of data can become so widespread that it may seem to be a fact. The so-called reward network is an example of this. Also, pre-existing understandings of love impact research strategies and interpretations of research data in ways that make it seem as if they are corroborated by CNS research. We have seen examples of this for the idea that mother love is biologically special and for the idea that love makes blind. In chapter 3 we will see some ways in which philosophers try to deal with interpretation in their research process. In chapter 4, I develop proposals on how interpretation in CNS could be improved with the help of those philosophical strategies. For now, I would like to conclude that particularly when cognitive neuroscience understands itself to be an interpretation-free research endeavor, it runs the risk of overlooking biases and other limitations in its operationalizations and interpretations of data.

The existential crises, questions and dissatisfaction that Anna, Bob, Clemens, Deidre and Edward experience do not only rely on their ability to love, however imperfectly, but also on their ability to self-reflect, however imperfectly. Self-reflection is a topic that cognitive neuroscientists investigate as well. In the next chapter, I review CNS of self-reflection along similar lines as I have reviewed CNS of love here. CNS of self-reflection is further developed than CNS of love: more studies have been performed and more critical attention has been paid to its operationalizations. Let us see what interpretative choices have been made there, what conceptual implications these have, and what potential contributions CNS of self-reflection has to make to our understanding of existential self-understanding.

2

Cognitive Neuroscience of Self-Reflection

Cognitive neuroscience (CNS) research into selfhood and self-reflection is approximately as young as CNS of love. The first study, entitled 'In search of the self: a positron emission tomography study' was reported in 1999, by Fergus Craik and colleagues. Unlike in CNS of love, many researchers involved in CNS of self are aware that operationalizing selfhood and self-reflection is fraught with difficulty. Attempts have been made to clear up conceptual muddles. Several involve a distinction between self-as-subject and self-as-object (Vogeley and Gallagher 2011; Christoff et al. 2011; Legrand and Ruby 2009). Most of the time, people are not consciously reflecting on who they are, but are simply going about the business of living. They perceive the world from their own perspective. They move around in the world, act and are acted upon. In these cases, people experience themselves as subjects, mostly implicitly. Their bodies and bodily perspectives distinguish them from others. This sense of self-as-subject and its neural correlates is examined e.g. in studies of the experience of agency, instructing participants to perform an action versus watch someone else perform that action (Farrer and Frith 2002) or instructing participants to imagine performing an action versus imagine someone else performing that action (Ruby and Decety 2001).

Self-as-subject is not specific to the existential aspect of selfhood or existential self-understanding. For current purposes, the other type of self-experience, which treats the self as an object of reflection, is more relevant. After all, the situations of existential dissatisfaction that Anna, Clemens, and Deidre find themselves in, move them to reflect explicitly on who they are. What is more, they could not end up in existential crisis, if it were not for their ability to reflect on their situation. Without their capacity for self-reflection, humans would not be able to form an understanding of themselves or to question themselves. In this chapter, I review CNS literature on self-reflection. As in the previous chapter, I examine the questions researchers aim to address with their studies, the choices they make in the way studies are set up and the conceptual implications of those interpretative choices for its potential contributions to human self-understanding. What types of insight to our understanding of existential selfhood does current CNS of self-reflection have to offer, particularly in light of the choices researchers make in their research process?

I performed Web of Science and PubMed searches with the key words 'self-reflection' and 'neuroimaging', read articles and review articles that I thus found and searched the reference lists of the articles I read for further CNS studies of self-reflection. For a study to be included, it had to satisfy the following inclusion criteria: it is a neuroimaging study; of explicit self-reflection; performed on a healthy population; and reported before 2015. This means I excluded studies that investigate neural correlates to self-reflection in patient populations, such as people with schizophrenia, autism, self-injury, etcetera, as self-reflection in those populations is presumably distorted. For similar reasons, I excluded a study with children as participants. I also excluded studies that do not investigate neural correlates to self-reflection per se, but rather the modulating influence of self-reflection on other cognitive processes, such as memory, attention, mind-wandering, and inferring other people's mental states. Studies of implicit self-experience or first-person perspective were excluded, including studies of agency. Likewise, I excluded studies of self-perception, such as studies investigating neural correlates to hearing or reading one's own name, or seeing one's own face, as they do not involve self-reflection. An additional search in November 2016 found four more studies. A total number of 24 + 4 studies were found. Details of these studies can be found in a table in appendix B.

As in chapter 1, this chapter also presents a conceptual review, of CNS of self-reflection. It proceeds somewhat more quickly, given that the research process and the moments of interpretation are virtually the same as in CNS of love and were already introduced in the previous chapter. First, I list the questions that current research in CNS of self-reflection aims to address; a list I have compiled based on the introduction and discussion sections of research articles. Second, I scrutinize the methodological choices researchers make in setting up their experiments and recruiting participants for them. Next, I give an impression of the types of results current CNS of self-reflection delivers. Then I draw out and scrutinize the conceptual assumptions regarding self-reflection that are embedded in current methodological choices. Finally, I draw conclusions about what self-reflection in fact amounts to in CNS of self-reflection and what this research can contribute to our understanding of existential self-understanding.

1. Research questions in CNS of self-reflection

All studies in CNS of self-reflection investigate what neural activity correlates with self-reflection. As in CNS of love, from the choice of contrast condition and the wider theoretical frame, I can infer researchers' interest in several sets of questions.

Researchers try to delineate the neural correlates of self-reflection from neural activity subserving closely related processes. They do so by means of introducing different control tasks.

- How does self-reflection differ from person-reflection more generally, in terms of its neural correlates? Specifically, how does reflection on oneself differ from reflection on a person to whom one is not close (Craik et al. 1999; D'Argembeau et al. 2005; Gutchess, Kensinger, and Schacter 2007; Jenkins and Mitchell 2011; Kelley et al. 2002; Kircher et al. 2000; Kjaer, Nowak, and Lou 2002; Modinos, Ormel, and Aleman 2009; Ochsner et al. 2005; Zhu et al. 2007)?
- How do the neural correlates to self-reflection compare to the neural correlates to reflection on a close other, such as a friend, sibling, mother or romantic partner (D'Argembeau et al. 2007; D'Argembeau et al. 2008; Kircher et al. 2000; Ochsner et al. 2005; Schmitz, Kawahara-Baccus, and Johnson 2004; Zhu et al. 2007)?
- How do neural correlates to self-reflection differ from or overlap with neural correlates to self-perception (Herwig et al. 2012)?
- To what extent do neural correlates to self-reflection regarding personality traits differ from or overlap with judging whether those personality traits are socially desirable (Craik et al. 1999; D'Argembeau et al. 2008, 2010; Ochsner et al. 2005; Schmitz, Kawahara-Baccus, and Johnson 2004)?
- Do neural correlates to reflection on people vary gradually with relationship intimacy (Ochsner et al. 2005; Murray, Schaer, and Debbané 2012)?

As in all cognitive neuroscience research, researchers would ideally like to find neural activity that is specifically dedicated to subserving self-reflection or one of its component processes. As research progresses, experiments are developed to *uncover more about the neural activity involved in self-reflection, and its components*.

- What do commonly found neural correlates to self-reflection tasks, such as activity in medial prefrontal cortex, stand for exactly? What may the functional components of self-reflection be with which neural correlates line up (Ochsner et al. 2005; D'Argembeau et al. 2007; D'Argembeau et al. 2008, 2010; Johnson et al. 2006)?

- Starting from an interest in self-reflection, for example: can cognitive and affective components of self-reflection be distinguished at the neural level (Moran et al. 2006)?
- And how do subcomponents interact with one another, e.g. where in the brain does self-related processing interact with emotional processing (Fossati et al. 2003)?
- Starting from an interest in the brain, for example: with what functional components of self-reflection does activity in anterior versus posterior medial regions line up (Johnson et al. 2006)? Is the insula reliably involved in self-reflection and can its activity be connected to a specific component of self-reflection (Modinos, Ormel, and Aleman 2009)?
- And how does medial prefrontal cortex develop functionally with age? Does its involvement in self-reflection differ for young versus old adults (Gutchess, Kensinger, and Schacter 2007)?

Researchers also try to delineate neural correlates to self-reflection from neural correlates to other types of self-experience, found in expected and somewhat unexpected places:

- How do neural correlates to self-reflection differ from or overlap with neural correlates to experiencing oneself as subject, in perspective taking (D'Argembeau et al. 2007)?
- How do neural correlates to self-reflection overlap with or differ from neural default activity, that is, neural activity when people are conscious but at rest as they are not instructed to do anything in particular (D'Argembeau et al. 2005; Gutchess, Kensinger, and Schacter 2007; Kjaer, Nowak, and Lou 2002; Macrae et al. 2004; Whitfield-Gabrieli et al. 2011)?

Researchers may be interested in a functional process and hope that neural activity *sheds light on an aspect of that functional process, especially where behavioral data are inconclusive* , e.g. because data cannot differentiate between different behavioral hypotheses. This is in fact how CNS of self-reflection came about:

- Why are people better at remembering information when it is related to themselves (self-reference memory effect)? Is processing (encoding) of information related to the self neurally dissociable from other types of semantic processing? What are the neural correlates underlying the self-reference memory effect: does the self have a privileged status in the brain (Craik et al. 1999; Heatherton et al. 2006; Kelley et al. 2002; Macrae et al. 2004)?

Several studies compare *different types of self-reflection* and the neural correlates associated with them. Altering the task through altering the instructions and/or stimuli allows for a comparison between different types of self-reflection to be made:

- Does reflection on your current self differ from reflection on your past self, in terms of the neural correlates involved? Is reflection on your past self more like self-reflection for your current self or like other-reflection (D'Argembeau et al. 2008)?
- Is reflecting on your past self similar to imagining your future self in terms of the neural activity subserving it? Is the difference with reflection on current self similar for both types of reflection on a temporally distant self (D'Argembeau et al. 2010)?
- Are the neural correlates to self-reflection the same for different aspects of the self? Specifically, is reflection on personality traits, current emotional/mental states and physical attributes subserved by similar neural mechanisms (Jenkins and Mitchell 2011)?
- Are neural correlates similar when reflecting on different aspects of the self even if these are presented in different ways, e.g. personality traits described in words and physical attributes depicted in pictures (Kircher et al. 2000)?
- Is self-reflection different in terms of its neural correlates when reflecting on what one generally is like versus what one is like in a specific context (Chiao et al. 2009a, 2009b)?
- Are neural correlates to self-reflection different when one reflects on what one is like directly versus more indirectly, by considering what other people (both close and non-close) may judge one to be like (D'Argembeau et al. 2007; Ochsner et al. 2005)?
- Can we find neural evidence for distinct self-knowledge systems, specifically intuition-based self-knowledge versus evidence-based self-knowledge (Lieberman, Jarcho, and Satpute 2004)?

Then there are a few studies in which participant selection influences the research question. These studies compare *different types of people reflecting on themselves*, under the assumption that neural activity correlated with self-reflection may differ for them, either because their brains are different or because they presumably understand themselves in fundamentally different ways:

- Does neural activity correlated with self-reflection differ for young adults versus old adults, i.e. for people with young versus old brains (Gutchess, Kensinger, and Schacter 2007)?

- Do neural correlates to self-reflection, reflection on a famous other to whom one is not close, and reflection on one's mother differ for Western versus Chinese participants? Specifically, is mother-reflection closer to self-reflection for Chinese than for Westerners (Zhu et al. 2007)?
- Does holding individualistic versus collectivistic values influence self-reflection and its neural correlates (Chiao et al. 2009a, 2009b)?

2. Operationalizing self-reflection

Researchers have to operationalize self-reflection to be able to study its neural correlates. They make choices regarding what participants to recruit. They also choose stimuli, develop instructions and choose a control task. Through these choices, studies pertain to something more specific than self-reflection. Here I summarize the choices researchers make. After giving an impression of the types of results current CNS of self-reflection delivers, I tease out the conceptual implications of these choices.

Participants

The people participating in CNS of self experiments are the people that generally participate in experiments: (undergraduate) students. Researchers make sure that participants are neurologically and psychologically healthy and do not use any drugs that affect brain activity. They select people who are right-handed. Other than these criteria to ensure that participants' brains and brain activity are similar enough, no special requirements are made on potential participants. Roughly equal numbers of men and women participate. In other words, unlike in CNS of love, participant selection does not seem to have conceptual consequences for what is actually investigated when investigating self-reflection.

Unless the fact that almost all participants are undergraduates living in WEIRD (Western, Educated, Industrialized, Rich, Democratic) societies (Henrich, Heine, and Norenzayan 2010) implies a particular self-concept rather than another. Zhu et al. (2007) think that this may be the case and compare brain activity of Western and Chinese students. They cite psychological research that suggests that Westerners have an independent sense of self, whereas Chinese have an interdependent sense of self, and hypothesize that this difference may be reflected on the neural level as well. Chiao and colleagues have compared American and Japanese students (2009b) for similar reasons and also conducted a study with bicultural Asian-Americans (2009a), priming their

collectivist or individualistic identities to see whether these are subserved by different neural activity patterns.

Gutchess, Kensinger and Schacter (2007) compare neural activity in young versus old adults, that is, in undergraduates (mean age: 23 years old) vs. people they call elderly (mean age: 72 years old). They are not primarily interested in potentially different self-concepts influencing neural correlates to self-reflection, however, but rather in how brains develop with age, and whether young versus old brains display functional differences in the medial prefrontal cortex.

Tasks: stimuli, instructions, control tasks

The majority of studies tapping into people's self-understanding use personality trait adjectives as stimuli. Examples of adjectives are: ambitious, patient, careless, stubborn, idealistic, arrogant, open-minded. Craik et al. (1999) were the first to do this, drawing their adjectives from Anderson (1968) and Kirby and Gardner (1972). They used 512 personality adjectives, divided into sixteen lists of 32 adjectives. They matched their lists for valence by ensuring that half of the adjectives on each list were positive and half were negative, such that at least in terms of social desirability, participants would be equally likely to identify with the personality traits on one list as on another. Craik and colleagues also made sure that each list contained roughly equal numbers of adjectives of two, three, four or five syllables. In that way, word processing for each list should be equally difficult. Several subsequent studies in turn derive their stimuli from Craik et al.'s study (Gutchess, Kensinger, and Schacter 2007; Ochsner et al. 2005), although many also select their adjectives directly from Anderson's lists (D'Argembeau et al. 2008, 2010; Fossati et al. 2003; Heatherton et al. 2006; Kelley et al. 2002; Kircher et al. 2000; Macrae et al. 2004; Moran et al. 2006; Zhu et al. 2007). Some come up with their own adjectives, use a different source, or do not mention their source (Chiao et al. 2009a, 2009b; D'Argembeau et al. 2007; Jenkins and Mitchell 2011; Lieberman, Jarcho, and Satpute 2004; Schmitz, Kawahara-Baccus, and Johnson 2004).

Self-reflection in these studies thus generally means self-reflection on one's personality traits, and researchers employ adjectives as stimuli. A few studies on self-reflection (also) choose other stimuli. Jenkins and Mitchell (2011) do not just include adjectives describing personality traits but also adjectives describing current mental states (e.g. bored) and stable physical attributes (e.g. tall). Kircher et al. (2000) do not just include personality trait adjectives for people to reflect upon, but also (modified) pictures of their faces. Two studies use short sentences describing all sorts of aspects about the self, not just personality: e.g. 'I get angry easily', 'my future is bright', 'I often forget important things'. Finally, in some studies participants are required to self-reflect

without encountering stimuli, about which more later (D'Argembeau et al. 2005; Herwig et al. 2012; Johnson et al. 2006; Kjaer, Nowak, and Lou 2002).

Most studies use personality trait adjectives as stimuli and require participants to judge whether a particular adjective applies to them. However, the particular *instructions* participants receive and the answers they are required to give come in different types. For example, Craik et al. (1999) ask their participants to judge how well they think an adjective describes them. The possible answers are 'almost never', 'rarely', 'sometimes', and 'almost always'. (Note how the distances between these response possibilities seem uneven: 'rarely' is closer to 'almost never' than 'sometimes' is to 'almost always'.) Similarly, Ochsner et al. (2005), in their second experiment, ask participants to judge the extent to which adjectives describe them on a scale from 1-4, where 1 equals 'not at all' and 4 stands for 'very much'. Although both studies pose the same question and give participants four response possibilities on a graded scale, Craik et al. and Ochsner et al.'s labelling of the response possibilities carry different conceptual assumptions about personality traits. The labels 'almost never', 'rarely', 'sometimes', and 'almost always' imply that the applicability of personality traits can vary from moment to moment (and therefore, one may infer, from setting to setting). On the other hand, a four point scale where 1 equals 'not at all' and 4 stands for 'very much' treats personality traits as varying in degree of intensity, keeping still about whether 'stubborn' describes people 'very much' as they are always very stubborn or stubborn at very many different occasions.

Most studies allow for two possible responses to a personality trait adjective: yes or no. Participants are instructed to judge whether an adjective describes them and then view a series of screens, saying, for example 'Myself (yes/no)? Honest' (Gutchess, Kensinger, and Schacter 2007) or 'Self + honest' (Zhu et al. 2007). Or, after receiving instructions, they simply view a series of adjectives, responding by pressing a yes button or no button to each of the adjectives (Kelley et al. 2002; Ochsner et al. 2005). Participants have to make up their minds about whether an adjective applies to them rather quickly. Studies allow participants anywhere between 2 seconds and 4 seconds to respond (Chiao et al. 2009a, 2009b; Craik et al. 1999; D'Argembeau et al. 2008, 2010; Gutchess, Kensinger, and Schacter 2007; Heatherton et al. 2006; Jenkins and Mitchell 2011; Johnson et al. 2002; Kelley et al. 2002; Kircher et al. 2000; Lieberman, Jarcho, and Satpute 2004; Macrae et al. 2004; Modinos, Ormel, and Aleman 2009; Moran et al. 2006; Ochsner et al. 2005; Schmitz, Kawahara-Baccus, and Johnson 2004; Zhu et al. 2007). That is to say, all studies using stimuli instruct participants to decide within 2-4 seconds, except for D'Argembeau and colleagues who allow 5 seconds response time (2007) and Fossati and colleagues who allow their participants 9.5 seconds (2003). The short response time in combination with yes/no answering possibilities imply that

participants access a rough and ready idea they have of themselves in these studies. The answer categories do not allow for subtleties along the lines of someone generally being 'honest' now, but less so in the past; of her being 'honest' under certain circumstances but not under others; of her being 'somewhat honest' overall; etcetera. There is no time for reflection on such subtleties either. Self-reflection in these experiments amounts to accessing a highly generalized understanding of one's personality.

Some authors explicitly acknowledge the generalized character of self-reflection in their studies, by phrasing their instruction as 'In general, I am patient' (Chiao et al. 2009a, 2009b; Jenkins and Mitchell 2011). Chiao and colleagues are interested to see whether accessing a rough and ready type of self-understanding differs from reflecting on oneself in the specific context of talking to one's mother: 'When talking to my mother, I am patient'. Jenkins and Mitchell, as said, compare reflection on a generalized understanding of personality to reflection on current mental states and physical attributes. Even the plain task of requiring participants to answer whether adjectives apply to them, yes or no, can be put to service in comparing different types of self-reflection. For example, Ochsner et al. (2005) wonder whether direct self-reflection is different from imagining how others would judge your personality and include conditions where participants are to judge the extent to which a friend or a neutral observer would deem personality traits to apply to them. Other researchers aim to access some of the variation in personality that may occur over time. They instruct their participants to reflect on who they are in comparison to who they have been or who they think they will be: 'At present, I am stubborn'; 'Five years ago, I was stubborn'; 'In five years from now, I will be stubborn' (D'Argembeau et al. 2008, 2010).

Four studies do not operationalize self-reflection as judging personality descriptions for their applicability to oneself. Instead, participants in these studies are required to self-reflect rather freely (D'Argembeau et al. 2005; Herwig et al. 2012; Johnson et al. 2006; Kjaer, Nowak, and Lou 2002). Instructions could run 'Try to think about your traits and attitudes, about your personality, in your relationship with others' for 100 seconds (D'Argembeau et al. 2005); or 'Think about your hopes and aspirations' for 18 seconds (Johnson et al. 2006); or 'reflect on who you are, how you are as a person' for 9.9 seconds (Herwig et al. 2012); or 'reflect on your personality' for 120 seconds, then 'reflect on your physical appearance', for another 120 seconds (Kjaer, Nowak, and Lou 2002). The time windows for self-reflection chosen here nicely illustrate how choices in operationalization are often driven by constraints stemming from imaging techniques. 9.9 seconds equals 5 full brain scans in these fMRI studies, just as 18 seconds equals approximately 9 full brain scans. The studies by D'Argembeau and colleagues and Kjaer, Nowak, and Lou make use of another imaging technique: Positron Emission Tomography, or PET for short. This technique requires longer time to obtain a full brain

scan, approximately 90 seconds, and 100 or 120 seconds of self-reflection thus accommodates a full brain scan.

After the experiment, researchers check whether participants have complied with the instructions through a free format interview. Participants report reflections such as "I am extrovert and talkative, while on the other hand I have sides only my friends know" (Kjaer, Nowak, and Lou 2002) and "I also thought that I am frank, if a friend does or says something that seems wrong to me I tell her or him immediately, and conversely, I like people to be frank with me and not to criticize me behind my back" (D'Argembeau et al. 2005).

As in all CNS experiments, the *control tasks* researchers use are of utmost importance for what they actually are testing. An early study compares self-reflection (i.e. judging personality trait adjectives) to rest (i.e. looking at a fixation cross) (Macrae et al. 2004). Any neural activity they find could simply designate processing of an adjective (as Gillihan and Farah note in their early review of control tasks (2005)). To eliminate the effects of simply processing an adjective, others include a control task such as judging whether an adjective is written in uppercase or lowercase letters (Gutchess, Kensinger, and Schacter 2007; Kelley et al. 2002; Zhu et al. 2007) or counting how many syllables it contains (Craik et al. 1999; Ochsner et al. 2005). Furthermore, neural activity found in the experiment should not be correlated with people-reflection generally, but rather with self-reflection specifically. Researchers therefore take activity correlated with self-reflection and subtract neural activity correlated with reflecting on someone known but not personally important, such as a president (D'Argembeau et al. 2005; Jenkins and Mitchell 2011; Kelley et al. 2002), queen (Kjaer, Nowak, and Lou 2002), prime-minister (Zhu et al. 2007), Einstein (Gutchess, Kensinger, and Schacter 2007), or an acquaintance (Modinos, Ormel, and Aleman 2009). Even then, however, the subtracted neural activity could signify reflection on someone who is personally important to one, rather than *self*-reflection per se. To ensure neural activity correlates with self-reflection only, participants should therefore also reflect on someone to whom they are close, such as a friend, sibling, romantic partner or their mother (D'Argembeau et al. 2008; Kircher et al. 2000; Ochsner et al. 2005; Schmitz, Kawahara-Baccus, and Johnson 2004; Zhu et al. 2007). This neural activity can in turn be subtracted from neural activity correlated with self-reflection. In that way, only neural activity specifically correlated with self-reflection remains.

3. Results and discussion in CNS of self-reflection

What types of results do experiments in CNS of self-reflection deliver? As often in CNS, main results in CNS of self-reflection consist of a list of neural correlates that activate specifically during self-reflection. Medial prefrontal cortex (MPFC) is virtually always activated, for example, and activity is also often found in other midline regions such as anterior cingulate cortex (ACC) and posterior cingulate cortex (PCC), extending into precuneus (for reviews, see Gillihan and Farah 2005; Legrand and Ruby 2009; Northoff et al. 2006; van der Meer et al. 2010).

In the discussion section, researchers interpret what activation of these cortical midline structures may stand for in functional terms. On the one hand, some researchers argue that cortical midline activation is specific to self-processing (Northoff et al. 2006). At the other end of the spectrum, some argue it has nothing to do with self-reflection per se, but indicates 'evaluation' more broadly: a combination of episodic memory retrieval and reasoning processes (Legrand and Ruby 2009). What is clear is that MPFC is also activated during reflection on other people (Legrand and Ruby 2009; van der Meer et al. 2010) and that MPFC and PCC also show activations during rest, that is, when participants lie awake in the scanner without having to perform any task (Whitfield-Gabrieli et al. 2011). Most discussion amongst researchers as well as further experiments aim to uncover whether overlap in activation patterns between self-reflection and closely related processes can be accounted for through overlap in functional components that these processes share. In that way, functional components of self-reflection may be specifically correlated to neural activity patterns.

For example, researchers consider the possibility that self-reflection consists of a cognitive component and an emotional component. Some argue that processing of the cognitive content may be subserved by dorsal MPFC whereas the emotional processing takes place in ventral MPFC (Northoff et al. 2006), whereas others link the cognitive component of judging self-relevance to the entire MPFC, and processing of emotional content to ventral ACC (Moran et al. 2006). As another example, researchers discuss the possibility that MPFC may become active during people-reflection in general, but the extent of the activation signifies the degree of self-relatedness, with more activity for close others than non-close others and most activity during self-reflection (D'Argembeau et al. 2010; Ochsner et al. 2005). As a last example, researchers discuss whether the overlap in neural activity between self-reflection and rest should be interpreted as indicating the general evaluative processes at work in both (Legrand and Ruby 2009) or whether it indicates that when at rest, people by default start to self-reflect (Gusnard et al. 2001). In sum, although certain activations are found consistently in CNS of self-reflection, their functional interpretation is still being debated and

investigated. Cognitive neuroscientists are therefore mainly working on the first two groups of research questions outlined previously, trying to find neural activity that specifically correlates with (components of) self-reflection. In other words, they are mainly investigating questions about the brain, not about the mind or about the behavior and experiences of people who self-reflect.

Several studies in CNS of self-reflection involve behavioral tasks and these can produce interesting results. For example, participants show a general tendency to deem positive traits more often descriptive of themselves than negative traits (e.g. Moran et al. 2006). And when past self, current self and future self are being reflected upon, participants are inclined to see their current self in more positive terms than their past self, and for their future self they embrace even more positive adjectives (D'Argembeau et al. 2010).

Chiao and colleagues (2009b) investigated self-reflection in Japanese and Americans. However, instead of assuming that native Japanese endorse collectivist values and Caucasian Americans endorse individualistic values, as many studies in cultural neuroscience do, they decided to test their Japanese and American participants beforehand, and to divide them into a collectivistic and individualistic group based on their scores. The collectivistic group ended up consisting of 5 Japanese and 9 Caucasian Americans, whereas the individualistic group consisted of 7 Japanese and 3 Caucasian Americans. There is nothing neuroscientific about these findings, but they provide important insight, particularly in light of other researchers' tendencies to simply assign participants to categories based on their continent of origin (e.g. Zhu et al. 2007).

Occasionally, behavioral results and the functional interpretation of neural results point in different directions. Heatherton and colleagues, for example, find themselves in a seeming stand-off between behavioral and neural data. Their behavioral data shows the self-reference memory effect: adjectives are recognized more accurately after being processed in terms of their descriptiveness of the self than after being processed in terms of whether they are written in uppercase or lowercase. However, adjectives are also recognized more accurately when judged for descriptiveness of a best friend, as compared to case judgments. Both these effects are big. The advantage of self-processing over friend-processing for memory performance is many times smaller. On the neural level, however, Heatherton and colleagues find a robust difference in MPFC activity between self-processing and friend-processing. If we are inclined to think this neural effect is representative of a behavioral effect too, then there appears to be a dilemma: should we believe people's brain or their behavior? This is a false dilemma, however. Heatherton and colleagues conclude that "although we may seem to incorporate knowledge about others into our self-concept by integrating them into our memorial

self-bias, the current results suggest that the neural mechanisms subserving this representation are likely to differ from the neural representation of self-knowledge" (2006). That is to say, they conclude that we have to reconsider our functional interpretation of this specific neural correlate, now that we find that behavioral data and neural activation patterns under our old interpretation of them, do not line up.

4. Conceptual implications: accessing a personality schema

As always in CNS research, the tasks employed in CNS of self-reflection have implications for the specific concept of self-reflection that is being investigated. This in turn has an impact on the contributions that current CNS of self-reflection can make to our understanding of the existential aspect of the human condition, as exemplified by the situations of Anna, Bob, Clemens, Deidre and Edward. We saw earlier that two broad classes of operationalizations of self-reflection can be distinguished. In the first type, participants are presented with words or short sentences and quickly have to decide whether these describe who they are. The second type of operationalization requires participants to freely self-reflect for a period of time. Both types tend to focus on personality rather than another aspect of the self, employing personality trait adjectives or instructing participants to freely self-reflect on their personality. Control tasks generally consist of having to judge descriptions regarding or reflect on someone else's personality. I would like to draw out the conceptual implications of these operationalizations now.

Self-*reflection* in these studies mainly amounts *to accessing a rough and ready representation* of the self. It is *rough* in that it is a schematized understanding, devoid of subtlety. As noticed above, there is generally no time nor an answering category that allows for subtleties along the lines of someone being 'stubborn' now, but less so in the past; of her being 'stubborn' under certain circumstances but not under others; of her being 'somewhat stubborn' overall; etcetera. That the required judgments indeed are more rough than the way in which people generally think about themselves becomes clear in the studies that ask participants to freely engage in self-reflection. These participants report thoughts such as "I am extrovert and talkative, while on the other hand I have sides only my friends know" (Kjaer, Nowak, and Lou 2002). Also, studies comparing reflection on general traits with reflection on traits in a particular context (calling one's mother) find that people judge themselves differently under these circumstances (Chiao et al. 2009a, 2009b).

Furthermore, reflection in most studies in CNS of self-reflection is conceived of as accessing a *ready* understanding of the self. In these experiments, participants do not reflect on themselves in order to gain a better understanding of themselves, but to draw on a general idea of themselves that they already hold. There is no time for anything else, and the experimental instructions induce this type of self-reflection. In any case, it stands in contrast to the self-reflection that Deidre engages in, for example. If she were to participate in an experiment of this type and had to decide in 3 seconds whether the term 'mathematician' describes her, yes or no, she would likely endorse it. She exactly engages in self-reflection, however, because the ready understanding that she has of herself, of Deidre as mathematician, does not feel right to her anymore, and she experiences a need to understand why and what an alternative to her current self-understanding may be.

A related point can be made about self-reflection as *accessing* a ready picture of the self. Operationalizations zoom in on self-reflection as akin to *passively* inspecting one's self-understanding. Whereas what often happens in real-life self-reflection is a construction of a self-understanding, in which the person reflecting plays an active role. Anna, for example, may not have reflected on herself much before her current bout of midlife crisis. If so, she is likely to develop a narrative understanding about herself in therapy at a level of detail that she has never developed before. Crucially, the understanding she articulates will differ somewhat depending on whether she is having a good day or a bad day, just as the understanding she articulates to her therapist is likely to be different from the understanding she articulates to her boss, were her boss to inquire why she is absent-minded and less productive than before. A study by Chiao and colleagues taps into this aspect of self-reflection through inducing either collectivist or individualist values in their bicultural participants and changing their self-reflections accordingly (Chiao et al. 2009a). Virtually all other studies, however, treat self-reflection as a passive reading off of a ready understanding of the self.

In the contrast between experimental task and the control task of judging or reflecting upon the personality of others, assumptions about selfhood are embedded. Selfhood is treated as something an individual is *at this moment, as compared to others*. 'At this moment' is one more way in which reflection revolves around accessing a rough understanding of the self. 'As compared to others' is not strange: we are in many ways the same as others, by virtue of being human beings, but selfhood is mostly conceived of as that which individuates us from each other. Nevertheless, this is not the only way in which selfhood can be operationalized. Different experimental and control tasks may embody an alternative conceptualization of selfhood, for instance, what an individual is like in relation to where she has come from and where she is headed.

Furthermore, the experimental situation as such induces self-reflection in an artificial, decontextualized manner. Deidre needs to take an important decision and this leads her to self-reflect. In different ways, Clemens and Anna are uncertain of their identities and they self-reflect because of this. Explicit self-reflection generally does not occur in a vacuum; people have reasons for self-reflecting and circumstances that make them self-reflect. In the experiments, however, *nothing is at stake* in self-reflection. Anna may break up her marriage and Deidre may switch careers depending on the outcomes of their self-reflections. More mundanely, people can engage in self-reflection when they find themselves experiencing things or acting in ways that they did not expect based on the understanding they hold of themselves. Self-reflection then leads to an update of self-understanding, even when it does not have any immediate consequences for behavior or life choices. In the CNS studies, however, there is no time to figure out something new about oneself. Nothing is at stake: quickly accessing a rough and ready self-understanding in a decontextualized experiment need not have any consequences.

Operationalizations of s*elf*-reflection can also be characterized by the aspect of the self that they focus on. In CNS of self-reflection this mainly is *personality*. And personality in turn is operationalized through a set of character trait descriptions. Participants in free self-reflection studies report reflections in which they also make use of personality trait adjectives ('I am frank'), with the difference that these endorsements of personality trait adjectives are mostly contextualized ('I am frank with my friends') and thereby moderated. Operationalizing reflection on oneself as reflection on one's personality is one way amongst others. However, Anna, Bob, Clemens, Deidre and Edward are not so much reflecting on their personality as on what is of utmost importance to them. They reflect on what they love. Clemens literally distinguishes this from personality by stating that he still has all his character traits, even though his identity is gone now that the family members he lived with and cared for have died.

Finally, the fact that data is obtained through averaging over participants means self-reflection is treated as a general process, independent of the particulars of the person doing the reflecting, both in terms of the self being reflected upon (each person having a different set of personality traits) and of the person doing the reflecting (and being better or worse at articulating nuanced self-understandings, more or less prone to self-serving biases and self-delusion).

To be clear, some researchers explicitly aim to activate rough and ready self-schemas in participants, especially when they are interested in the self-reference effect on memory, more than in self-reflection per se. This is the case for several early studies and these do not use the label 'self-reflection' much either. As operationalization is difficult and reviewers always like references to other studies that use the same operationalization, later studies have copied this operationalization in the context of

their interest in the neural correlates to self-reflection. Naturally, one has to start somewhere and the operationalization of judging personality adjectives taps into some aspect of self-reflection. However, any operationalization has conceptual implications regarding the process under study. It is vital to be aware of these implications. Trouble can easily ensue when a task is transplanted from one context (e.g. aiming to activate a rough self-schema) to another (e.g. aiming to induce self-reflection in participants).

This brings us to a final point: the issue of labeling. Take D'Argembeau and colleagues, who have contributed four studies to the literature. What these studies purport to be measuring is labelled as neural correlates to 'self-referential reflective activity' (2005), 'self-referential processing' (2007), 'self-reflection' (2008), and 'thinking about selves' (2010). The first label refers to the task involving free self-reflection for 100 seconds, whereas the other three labels refer to the task of quickly judging personality adjectives on whether they are descriptive of oneself. Several other labels also occur in the literature, such as neural correlates to 'self-knowledge' (Lieberman, Jarcho, and Satpute 2004; Ochsner et al. 2005), 'self-awareness' (Kjaer, Nowak, and Lou 2002), 'metacognitive evaluation of the self' (Schmitz, Kawahara-Baccus, and Johnson 2004), 'judgments of self' (Jenkins, Macrae, and Mitchell 2008), or simply 'the self' (Fossati et al. 2003; Heatherton et al. 2006). Besides these different labels, researchers also have different assumptions as to what psychological category they are investigating the neural correlates to: some speak of neural correlates to the *process* of self-reflection, others of neural correlates of a *representation* of the self, and yet others of the neural correlates to self-*experience*. Different labels complicate the compilation of databases and review articles. Over and above that, assumptions regarding the psychological category whose neural correlates are being investigated point into different research directions and conceptualizations of components. Again, the importance of the scrutiny of conceptual assumptions cannot be overemphasized. In chapter 4, I delve into these issues into more detail when I discuss my ideas about how philosophy may be of help to cognitive neuroscience.

5. Conclusions CNS of self-reflection

The human capacity for self-reflection is an important contributing factor to human self-understanding and cognitive neuroscientists study its neural underpinnings. What types of insight to our understanding of the existential aspect of the human condition does current CNS of self-reflection have to offer? In this chapter, I have reviewed the choices

researchers make in their research process and the conceptual implications of those choices for the types of questions experiments address.

As in all CNS research, interpretation also plays a role in the research process of CNS of self-reflection. Selection of participants may not have conceptual implications to the extent that it has in CNS of love, but operationalization of self-reflection into a task involving stimuli and instructions, plus a control task to compare it to, most definitely has. Unlike in CNS of love, participants have to actively perform a task and we saw that response categories and response time have conceptual consequences. Self-reflection is operationalized in broadly two ways: through a task of participants being faced with a series of personality trait adjectives and quickly having to judge whether they describe themselves; and through a task of participants having to self-reflect freely on their personality. In both tasks, the aspect of the self that participants reflect on is personality, not the existential aspect to selfhood. In both tasks, also, the experimental situation induces self-reflection without anything being at stake, unlike most natural cases of self-reflection. Besides that, self-reflection in the second task is much like self-reflection in the wild, so to speak, that is, self-reflection like Anna or Edward engages in. Self-reflection in the first task, on the other hand, is conceived of as passively reading off a rough and ready representation of the self. This is unlike the self-reflection we encountered in the vignettes in the introduction: there, people self-reflect because their rough and ready ideas about themselves do not suffice in one way or the other.

Critical reflection on interpretative moments in the research process and their conceptual implications is crucial, first of all, to safeguard the validity of operationalizations and the validity of conclusions drawn on the basis of research. Second, it is also crucial to ensure the comparability of studies to each other and thereby to ensure the quality of review articles and databases. Labels used in CNS of self-reflection vary widely. Different labels are often used to refer to the same task, however, and studies employing this task should be categorized together when aggregating data. Also, different researchers may use the same label to refer to the two different types of tasks used in CNS of self-reflection. If all studies reviewed here are included in the same category, the difference between these two tasks is glanced over. Reviews and databases of neural results thus become muddled (about which more in chapter 4). This chapter may serve as conceptual review of CNS of self-reflection; a necessary precursor I would say to setting up informative reviews and databases of neural results. Finally, conceptual review is also an excellent way to determine how to move CNS of self-reflection forward. It would be interesting to see whether current patterns of results replicate in more studies employing free self-reflection operationalizations. If researchers are interested in looking at self-reflection and its neural correlates from different sides, they may try

to investigate self-reflection on other aspects of the self besides personality, such as the existential aspect.

Neural activity correlating with self-reflection is rather consistently found in medial prefrontal cortex (MPFC) and other midline regions. As of yet, it remains unclear what aspect of self-reflection tasks this activity correlates with exactly. Other tasks besides self-reflection tasks also correlate with MPFC activity, after all. That is to say, even when a consistent neural correlate is found in a set of studies, its functional interpretation remains unclear, as that neural activity virtually always also correlates with other processes. Research efforts therefore continue to focus on trying to delineate what functional component correlates exactly with what neural activity. The contribution to self-understanding that CNS of self-reflection promises to give therefore concerns insight into the neural underpinnings of the human capacity to self-reflect. As a source of insight into self-reflection, it will always be indirect.

Biases are less prominent in CNS of self-reflection than they are in CNS of love. One study compares neural activity in Chinese and American participants, under the assumption that they have different self-concepts (interdependent vs. independent, respectively) and are therefore bound to self-reflect differently. When differences in the neural activity patterns of Chinese and American participants are indeed found, researchers conclude that their study "provid[es] neuroimaging evidence that the 'Western self' is different from the Chinese self at a neural level" (Zhu et al. 2007, 1315). Although Zhu and colleagues strictly speaking only refer to finding neuroimaging evidence for a difference 'at a neural level', it is easy to mistake their findings as evidence on the neural level for a difference at the psychological level, between Eastern interdependent self versus Western independent self. The effects they found may indeed reflect this, but they may also reflect differences in the testing conditions between China and the US, or differences in the reflection strategies of the Chinese and the Americans that do not have to do with their self-concepts per se, etcetera. In any case, these results are complicated considerably by studies that do not assume Easterners to have interdependent self-concepts and Westerners to have independent ones, but test these assumptions beforehand and find rather more complex patterns of results (Chiao et al. 2009a, 2009b).

Which brings us to the final point of interest of this chapter and it is one that did not feature in CNS of love. In the process of participating in CNS experiments on self-reflection, participants often have to complete a behavioral task too. These tasks are more likely to produce results that challenge prevalent understandings of ourselves and other people than the neuroimaging component to the experiment. The study by Chiao and colleagues is a good example of this. They put prevalent ideas about Eastern collectivistic (or interdependent) self versus Western individualistic (or independent)

self to the test, simply for the sake of dividing up participants along collectivistic versus individualistic lines, i.e. before their actual experiment even started. It turned out that 9 Americans and 5 Japanese had a sense of self best characterized as collectivistic whereas 3 Americans and 7 Japanese had an individualistic sense of self. With respect to our understanding of people, this result is more interesting than the fact that they found MPFC activity correlated with their personality trait judgment task. What is more, behavioral tasks and neural activity patterns may not line up in terms of what they suggest about behavior and experiences. As long as we are uncertain what neural activity stands for in functional terms, it seems prudent to assume the behavioral results give us the best ideas about behavior and the interpretation of what neural activity patterns reveal about behavior needs to be fine-tuned further.

We will return to many of these issues in chapter 4, when I discuss what a conceptually improved version of CNS may look like, as well as what this CNS of the future may have to contribute to our understanding of existential self-understanding. First it is time to turn to philosophy, however. I want to examine philosophy's resources for understanding the existential aspect of the human condition as it is lived by Anna, Bob, Clemens, Deidre and Edward. What has philosophy to offer to our understanding of self-understanding and what role does interpretation play in its research process, exactly?

3

Philosophers of Existential Selfhood

Anna feels disconnected from her marriage and her work. She doubts her love for her husband. She experiences meaninglessness and dissatisfaction with her entire life and, eventually, seeks out therapy for it. Bob has a relation to his dog that impacts his behavior and experiences. This relationship is troubled, however; it is an attachment characterized by negative feelings and hurtful behavior. When the dog disappears, Bob is torn as to whether to try and find his dog or to leave it be. Clemens had identity and thereby direction, purpose, meaning in his life, through his relations to his wife and children. As these have fallen away, he does not know who he is anymore and needs to find this out anew. Deidre devotes much of her time to mathematics and thinks of herself as a mathematician. Yet she feels less and less connected to her work. The philosophy books and lectures she encounters attract her much more. Over the course of a few years, she reappraises her professional identity. Edward's understanding of himself is entirely tied up with him being a butler. He has always acted according to the requirements of his profession, as he saw them. Approaching old age, he comes to realize that he may have been mistaken to do so and that his self-understanding is flawed in many ways. It does not include, for example, a proper appreciation of his love for Ms. Kenton, the housekeeper.

How should we understand the existential aspect of the human condition? We saw that CNS research into love and into self-reflection first and foremost aims to investigate the brain, particularly the neural activity that correlates specifically with (components of) love and self-reflection. Although the relevance of this research is often couched in terms of self-understanding, it can only contribute indirectly to our understanding of behavior and experience, through inferences drawn from neural data. As of yet, it has not much insight to offer regarding the human propensity for existential self-understanding. How about philosophy? What can current philosophy contribute to our understanding of existential selfhood and existential self-understanding? And what role does interpretation play in its research process? Is philosophy really a speculative prelude to experimental science, as some would have it (Churchland 2008)?

Philosophers speak of 'the self' to indicate they do not speak about anybody's self in particular, but about the common structure underlying people's different selves. Most

philosophers I consider in this chapter give an account of the self as it comes to the fore in existential situations. All of them aim to elucidate existential phenomena: meaning in life; alienation; existential questions and dilemmas; selfhood in terms of what is of utmost concern to oneself and distinguishes oneself from others along these lines; self-understanding as an understanding thereof. Susan Wolf and Harry Frankfurt both examine the role love plays in people's lives. Second, narrativists focus on the stories people tell or at least are able to tell about themselves and that shape their lives accordingly. Marya Schechtman is arguably the most important narrativist who is still working, while Charles Taylor's thought is relevant as he has tried to articulate the ways in which narratives—and thereby selves—revolve around what is of fundamental importance to people. Finally, Søren Kierkegaard considers existential selfhood in terms of the relation in which we stand to ourselves. He is included as one of the fathers of existential philosophy. His thought has recently been brought to bear on contemporary debates surrounding narrativist theories of the self (Davenport and Rudd 2001; Lippitt 2007; Rudd 2007, 2012; Stokes 2010) and Frankfurt's thought on existential selfhood (Rudd and Davenport 2015; van Stee 2015b). [3]

In the first three sections of this chapter, I examine the contributions these philosophers make to our understanding of existential selfhood. Throughout, I evaluate the extent to which their views elucidate the existential phenomena at stake in the situations of Anna, Bob, Clemens, Deidre and Edward: how we are shaped by loving others (Wolf, Frankfurt); by the narratives we hold about ourselves (Schechtman, Taylor); and by the relation in which we stand to ourselves (Kierkegaard). In chapter 5, I aim to draw out these lessons on existential selfhood and combine them, improving on their shortcomings where necessary. In the final two sections of this chapter, I take a step back and take a (metaphilosophical, if you wish) look at the types of questions philosophers address and, second, the role interpretation plays in their research process. I also compare philosophy to cognitive neuroscience on these two points.

1. Being a self by loving others

"When I visit my brother in the hospital, or help my friend move, or stay up all night sewing my daughter a Halloween costume, I act neither for egoistic reasons nor for

[3] Christine Korsgaard's work on self-constitution focuses on the common structures that allow all of us to be agents, and moral agents at that, in spite of our different practical identities (2009). I do not discuss her views here, nor any other views that revolve primarily around moral agency, instead of around existential motivation and existential experience irrespective of morality.

moral ones. [...] Rather, I act out of love" Susan Wolf writes (2010, 4). Both she and Harry Frankfurt work within practical philosophy, and find its ongoing conversation on the human condition wanting. Practical philosophy and its accounts of people's sources of motivation and reason tend to focus on morality and egoism. People's drive towards leading a life they can think of as meaningful receives insufficient attention in practical philosophy according to these philosophers, just as the experience of alienation from one's life does not feature in their colleagues' accounts of the human condition. Wolf and Frankfurt articulate and think through what I would call the existential aspect of the human condition and thus aim to convince their fellow practical philosophers of the need to do so.

Meaning in life through loving what is worthy of love

The questions Wolf aims to address mainly concern the nature of meaningfulness and its connection to love. Wolf writes: "According to the conception of meaningfulness that I wish to propose, meaning arises from loving objects worthy of love and engaging with them in a positive way" (2010, 8). She fleshes out the conditions under which this applies. First, Wolf introduces a subjective criterion: experiencing love for something, at least usually. "A person's life can be meaningful only if she cares fairly deeply about some thing or things, only if she is gripped, excited, interested, engaged [...] as opposed to being bored by or alienated from most or all that she does" (2010, 9). Second, this subjective criterion needs to be paired with an objective criterion according to Wolf, for feeling good about what you do does not guarantee it contributes to meaningfulness in your life. Whatever it is you love has to be worthy of your love. "A person who loves smoking pot all day long, or doing endless crossword puzzles, and has the luxury of being able to indulge in this without restraint does not thereby make her life meaningful" (2010, 9). Lastly, there has to be an active type of relationship between the person and her object of love. She has to be able "to protect it, promote it, honor it, or more generally, to actively affirm it in one way or another" (2010, 10).

Wolf thus tries to articulate the conditions that are individually necessary and collectively sufficient for a life to be considered as meaningful. They can be used to understand some of the general characteristics of the situations Anna, Bob, Clemens, Deidre and Edward find themselves in. For example, Anna and Deidre no longer feel love for the person and/or pursuit they devote much of their life to, corresponding to Wolf's subjective criterion. The objective criterion is at stake when Edward comes to doubt the objective worth of his life's work when he has to acknowledge that he has loyally served a lord who sympathized with the Nazis. It may also be at play in Bob's situation, devoting his life to a dog as he does. Furthermore, Bob cannot really be said

to actively protect and honor the dog he devotes his time and energy to; on the contrary, he kicks it and swears at it, violating the final criterion. Wolf's views on meaningfulness thus allow us to understand why, when reading about these people's lives, we judge something to be problematic in each of them and to understand more specifically *what* it is that is problematic. It also enlightens us on the kinds of questions people may ask themselves when striving for existential self-understanding, particularly with a view of making their lives more meaningful. According to Wolf, they should ask themselves: what am I truly enthusiastic about? Is what I am devoted to worthy of my devotion? And do I positively contribute to what I love, or actively affirm it in any other way?

On the other hand, we may also read e.g. Bob's situation differently and question Wolf's account accordingly. Maybe Bob has mental issues and caring for a dog in the way that he does may be the best he can do. Who are we to judge his life as meaningless? What does the objective criterion amount to exactly? Both Wolf and her critics (e.g. Frankfurt 2002; Arpaly 2010) discuss the objective criterion at length. Wolf considers the possibility that the worthiness of projects may stem from their value lying, at least partly, outside of oneself (2010, 41). Alternatively or moreover, the objectivity constraint may require that judgments of value are subject-independent, such that one can be mistaken in thinking that a project can be a potential provider of meaning to life (2010, 43). In the end, Wolf confesses that she is not completely satisfied with any of these accounts of the objectivity of values (2010, 45), but holds firm with her argument that "value judgments are not radically subjective" (2010, 47). She discusses a witty range of examples (Sudoku solving, pole sitting and goldfish watching are amongst them) as not providing meaning to life. Wolf also readily acknowledges that she might be wrong about the particular examples and that in fact no one can be said to be the final authority on the objective value of things. This "does not call into doubt the legitimacy or coherence of the question itself [of which things have value] or of the enterprise of trying to find a more or less reasonable, if also partial, tentative, and impermanent answer" (2010, 40). In other words, Wolf stresses the importance of the question even if her answer is tentative: when people search for existential self-understanding, they are also interested in determining whether what they are devoting their lives to is worthy of their devotion.

As said, Wolf's aim is to point out the relevance of meaningfulness and love to practical philosophers who mainly focus on morality, and elaborate a view on their relation. Wolf does not develop a view on how what we love and what contributes to the meaningfulness of our life may be constitutive of (the existential aspect of) the self. She occasionally implies that what we love matters for who we are, for example when she speaks of "[t]he role of meaning in a person's life, and the character of a person's attachment to the things that give her life meaning" (2010, 54) or when she identifies alienation as an experience diametrically opposed to the experience of meaning in life

(2010, 112). Unpacking the relation between love and meaning in life is the aim she has set herself, not the relation between love, meaning and selfhood. This is done by Harry Frankfurt, whose research interests and use of the concepts of 'love' and 'reasons of love' overlap to quite an extent with Wolf's (Wolf 2010, 4n, 2002; Frankfurt 2002).

Love structuring selfhood

In his own words, Frankfurt focuses on "various interrelated topics in moral philosophy and on what I suppose may plausibly be construed as philosophical anthropology" (1999b, ix). He approaches philosophical anthropology as follows: "In general, the approach I take in trying to understand what we are is to consider the structure and constitution of the self. My emphasis in this is mainly on the will" (1988, viii). In Frankfurt's view on the constitution of the self, love plays a pivotal role. "Our essential natures as individuals are constituted [...] by what we cannot help caring about. The necessities of love, and their relative order or intensity, define our volitional boundaries. They mark our volitional limits, and thus they delineate our shapes as persons" (1999a, 138). Frankfurt's work is mainly relevant towards understanding the role love plays in existential selfhood and existential self-understanding with respect to two questions. First, how should we understand love? Second, how does love shape who we are? That is, how should we understand the constitutive role love plays in existential selfhood?

Frankfurt describes love as a "concern for the existence and the good of what is loved" (2006, 40, 2004, 42). In his view, this concern is rigidly focused, non-utilitarian, involuntary, and constitutive of the self (2006, 40–43, 2004, 41–47, 79–80). He unpacks this as follows. Love is *rigidly focused* in that we love a particular individual instead of loving the individual as representative of a larger class of people. Love is thus different from most charity, where it does not matter what individual is to be helped, as long as he is sick, poor, or a refugee (2004, 43). Clemens cannot just forget about his deceased wife and children and marry any other woman. Even if, say, a widow exists who has children of approximately the same age as Clemens' deceased children and who shares other characteristics with Clemens' deceased wife (her even character, her profession, the region she grew up in) he cannot simply substitute his deceased family with this other family. His love is rigidly focused on the particular people that his wife and children were.

Furthermore, Frankfurt deems love to be *non-utilitarian*. When we love someone or something, it functions for us as an end in itself, not as a means to some other end. We may only care about the well-being of an employee because he cannot function professionally without it. And our care for social justice may be motivated by our

profound dislike of riots. However, when we love the employee or when we love social justice, we treat them as final ends, not as a means towards production or peace and quiet, respectively (2004, 42). Thinking along with Frankfurt, we can imagine Anna to have been devoted to her work and her husband as ends in themselves, in previous, happier times in her life. At the same time however, contra Frankfurt, actual love relationships do not appear to ever be entirely non-utilitarian. Even if we do not primarily treat our objects of love as means to other ends, they tend to fulfill instrumental roles in our lives as well. Anna's work enables her to pay the bills, for example. And her husband is capable of cooking her meals, taking out the garbage and, not in the least, to take away some of her loneliness. There are bound to be moments at which these instrumental functions of work and husband get meshed up with or take priority over her loving concern for them and their interests as ends in themselves.

The *involuntariness* of love is mostly described by Frankfurt as the necessity of love. Loves are volitional necessities: it is not up to us to determine what we love. What is more, once we love someone or something, we cannot but devote ourselves to promoting the interests of the beloved. In these ways then, a lover is not free. "On the contrary, he is in the very nature of the case *captivated* by his beloved and by his love. The will of the lover is rigorously constrained. Love is not a matter of choice" (1999a, 135, his emphasis). Frankfurt mentions that we sometimes try to get rid of our love for someone, trying to change the circumstances such that we may love someone else instead or at least stop loving what we do not wish to love. Mostly, however, we cannot bring ourselves to even try and in any case, such attempts can only succeed up to a certain extent. In the end, "[w]hat we love is not up to us" (2004, 49).

That we have little to no control over what we end up loving does not mean that loving makes us entirely unfree. Quite the contrary, according to Frankfurt. Even though we are not free to determine what we love and thus are not free to determine how our existential selfhood, or will, is shaped, when we act according to what we love, we are entirely free and experience this too. At such moments, we act according to what is most important to us. What is more, loving makes it easier for us to make choices. Thus, the binding of our wills by love can be experienced as liberation (1999b, x, 1999d, 2004, 64).

Another important consequence of Frankfurt's insistence on the involuntariness of love is that it ensures a measure of continuity in the existential aspect of our selves. We cannot decide on a whim that we will start loving someone or something or will stop doing so. Instead, love happens to us. We discover what we love, thus discovering who we are, existentially speaking. We cannot change ourselves at will. The existential self is constituted by stable volitional tendencies. Do not be confused by the terminology here: saying that these volitional tendencies are involuntary is not a contradiction. It is the same as saying that we cannot alter the structure of our will at will. Our most

fundamental (volitional) motivations stem from love, yet we do not voluntarily decide what we love.

The involuntariness or necessity of love explains why Bob cannot just decide to forget about his dog and move on with life. It also makes sense of why Deidre cannot rekindle her love for mathematics at will and why she experiences her attraction to philosophy as a discovery rather than a decision. At the same time, however, people try to influence what they love and how they relate to what they love. Anna seeking out therapy is a good example. As said, Frankfurt notes this possibility (2004, 49), but does not include it in his analysis of love as a concept, nor in his analysis of the structural relation between love and selfhood, as we soon will see.

Finally then, love is *constitutive of selfhood*, according to Frankfurt. This brings us to the second question regarding existential selfhood that Frankfurt addresses: how should we understand the structural relation between love and the existential aspect of the self? Frankfurt writes:

> [A] lover *identifies himself* with what he loves. In virtue of this identification, protecting the interests of his beloved is necessarily among the lover's own interests. The interests of his beloved are not actually *other* than his at all. They are his interests too. Far from being austerely detached from the fortunes of what he loves, he is personally affected by them. [...] The lover is *invested* in his beloved: he profits by its successes, and its failures cause him to suffer. To the extent that he invests himself in what he loves, and in that way identifies with it, its interests are identical with his own (2004, 61–62).

What we love affects who we are. That this is so can be seen according to Frankfurt in love's influence on both our experiences and our actions. Our loved ones do not leave us cold: we are personally affected by the fortunes of what we love. In Frankfurt's view, there is a direct, positive connection between the lot of what we love and our experiences: when our beloveds are doing well, we prosper because of it, just as when things do not go well for them, we suffer too. There is thus little distance between our beloveds and ourselves. Where the interests of our beloved are concerned, this distance disappears entirely: lovers identify with what they love, Frankfurt writes, thus including the interests of their beloved people and projects among their own interests. Our beloved's interests are identical with our own.

The structural relation between love and the self is thus conceptualized by Frankfurt as *identification*. More particularly, as Frankfurt always focuses on the will when thinking about the constitution of the self, it is one of *volitional identification*.

Frankfurt conceives of loves as volitional necessities, constitutive of selfhood as we volitionally identify with what we love. Loves are inherent to the will and structure the will. Loving thus gives us reasons for action: we have reason to act in order to protect and promote the interests of our beloved people and projects, as their interests are included amongst our own, as our own. Although loving is often accompanied by positive feelings and attitudes towards the beloved, these are not necessary conditions of love. Frankfurt writes that "enthusiasms are not essential. Nor is it essential that a person likes what he loves. He may even find it distasteful. As in other modes of caring, the heart of the matter is neither affective nor cognitive. It is volitional" (2004, 42; see also 1999a, 129; 1999c, 161).

I would like to question Frankfurt's characterization of the structural relation between loving and the existential aspect of ourselves as volitional identification in several ways. First, 'identification' takes our care for our beloveds and their interests rather far. We do not coincide with the people and projects that we love; there always remains a distance between us and them that Frankfurt's use of 'identification' does not seem to account for. Bob, for example, seems far from accepting the interests of his dog as his own. He may walk the dog according to its needs, but at the same time kicks the dog, much against the interests of the dog. His whole relationship with his dog seems characterized by such ambivalence and his practical identity is likewise ambivalent: when the dog disappears, he cannot decide whether to go and look for it or not. He is angry at the dog for disappearing and sad for the dog's disappearance at the same time. Given that, according to Frankfurt, we identify with our beloveds and their interests, such ambivalence in our relations to our beloveds and the accompanying ambivalences in ourselves, cannot be accounted for in his terms. The sources of ambivalence that Frankfurt mentions have to do with the interests of two loves pointing in different directions, or with a beloved's interests clashing with other interests that the lover has (2004, 62). That ambivalence could also result from an ambivalent relation to a loved one is not on his radar. Such ambivalences cannot find a place in the structural relations between us and what we love, as identification takes away the distance between us and them. (For a more extensive discussion of this issue, and rebuttals of alternative readings of the relationship between Bob and his dog, see Van Stee (2015a).)

To put it differently, Frankfurt's view of selfhood as constituted by our identification with what we love dissolves the self too much in that it surrenders the self to others and their interests. Edward may be said to identify, in Frankfurt's sense, with his work as a butler at Darlington Hall, as he fully accepts the interests of that line of work as his own. Yet he thereby does not make use of a human capacity many human beings would put to work in his situation. Edward never questions his devotion to the

lord he serves, even when that lord fires two Jewish employees and invites Nazis over for tea. His life can virtually entirely be subsumed under his butlership; he never questions whether he has other interests and whether these should not occasionally trump the interests of his line of work. Edward does not distance himself from his line of work and appears to be an impaired self because of it.

Then there is Frankfurt's view that love is first and foremost volitional and can be divorced from positive affect. This cannot be the standard case, not even for Frankfurt: when he writes that our identification with our beloveds means that we profit by their successes and suffer when they fail, it is hard to see how this could be understood without reference to affect. Frankfurt appears to think the case in which we do not like what we devote ourselves to is unquestionably a case of loving and a case in which our volitional identification with what we thus care for, even though we do not like it, constitutes us. Yet it seems to me that when we do not like what we devote ourselves to practically, this cannot but be a problem for us. It puts us at risk of self-alienation. Deidre, for example, can be said not to like mathematics very much anymore, even when she still devotes her days to it. In Frankfurt's terminology, Deidre would still appear to love mathematics, as she still acts out of a practical concern for it. If she were to continue to do so, however, in spite of liking mathematics less and less, she is most certainly going to experience "a kind of nagging anxiety or unease", feeling "troubled, restless and dissatisfied with [herself]", that is, she is going to be prone to the psychic distress that Frankfurt diagnoses as the opposite of "feeling at home with ourselves" (2004, 5) that loving and acting out of love should provide. If Frankfurt wants to hold on to the distinction between affect and volition in loving, he cannot elucidate experiences of meaningfulness and meaninglessness anymore. If he wants to clarify such experiences, he should connect affect and volition with each other, as Wolf does with her subjective criterion.

Like with Wolf, we can understand Frankfurt's motives for developing his ideas about how love shapes the existential aspect of selfhood in the way that he does if we consider the overarching aims he has set himself. Given that Frankfurt aims to convince fellow practical philosophers of the need to take love into account, his emphasis on the *volitional* nature of love makes some sense. Frankfurt wants to argue that love does not just motivate people to act so as to benefit what they love, he also wants to argue that people are justified in doing so: loving gives people reason to act according to the interests of their beloveds. Frankfurt likely wants to avoid building his account of practical reason on feelings, as they may be fleeting and occur coincidentally and therefore lack normative force.

Furthermore, Frankfurt's account of love constituting selfhood through volitional *identification* with beloved people and projects and their interests ensures a measure of continuity with his early work on freedom of the will. There, Frankfurt distinguishes between desires that arise in us (we want ice cream) but that we distance ourselves from (we do not want to want ice cream) and desires that arise in us that we identify with (we want to want carrots). Desires that we identify with are internal to our wills and thus part of ourselves. When we act on such desires, we act freely as *we* are the ones that act. Conversely, desires that arise in us but that we distance ourselves from are external to our wills and thus ourselves; when we act on them nonetheless, we do not act freely (Frankfurt 1971, 1999c, 159). Understanding our relation to our beloveds as one of necessary volitional identification allows Frankfurt to build on his earlier thought and argue that what we love is necessarily internal to our wills and thereby ourselves. He thus ensures that when we act based upon what we love, we act freely.

Nevertheless, it is something rather different to identify with desires arising *within us*, as per the early work, or identifying with what we love, which is *outside of us* (Kroeker 2011), particularly when beloveds outside of us are autonomous people. (This issue may easily go under the radar of Frankfurt's readers as identifying with people we love sounds natural in everyday English, more natural in fact that identifying with our desires.) The problems this may lead to may be obscured somewhat by Frankfurt's idealization of love as non-instrumental. After all, little harm seems done if we identify with our beloveds and their interests insofar as these beloveds also love us, and non-instrumentally so. Yet our beloveds are only human and they may not care as much about us as we care about them, or they may care more about what we can do for them than what they can do for us. As Wolf said, we may question whether what we love is worthy of our love. Also, I would add, we might ask ourselves how we should deal with and relate to the people that we love. In Frankfurt's account there is no space where such questions can arise as 'identification' with our beloveds, i.e. our acceptance of their interests as our own, removes the distance between us and them. Ironically, a term that was originally developed to help us understand the freedom of the will that we as humans have, is now used in such a way that some of our freedom disappears from view.

Love, self and philosophy

All in all then, we see that these philosophers of existential selfhood, Susan Wolf and Harry Frankfurt, endeavor to delineate necessary and sufficient conditions for love and meaning in life, as opposed to experiences of meaninglessness or alienation from one's life. Frankfurt in particular also tries to account for the structural relations between love and existential selfhood. While cognitive neuroscientists contribute an understanding

of the brain's involvement in the capacities enabling existential self-understanding (i.e. the capacities to love and self-reflect), these philosophers clarify the structural characteristics of the existential phenomena exemplified by Anna, Bob, Clemens, Deidre and Edward.

In the process of aiming to say something about meaningfulness in life and the relation between love and selfhood, respectively, Wolf and Frankfurt come across structural connections to other phenomena than the ones we were originally interested in. Frankfurt helps us understand the basis for a continuity in the existential aspect of the ourselves, for example: as we cannot start or stop loving at will, we cannot easily change ourselves in terms of what we love. Also, in the process of speaking about how loving gives us reasons for action, Frankfurt clarifies how it can be that loving is experienced as both captivating and liberating: we cannot help what we love and are unfree in this sense, but loving provides us with our most fundamental motives for acting and thereby liberates us from the potentially crippling uncertainty as to what to do.

If Anna and the others want to resolve the situations they are in, they need to take into account the particularities of their lives and their loves. Philosophers do not deal with such personal particularities; they are not therapists. Philosophers work at a more abstract level, aiming to articulate common characteristics underlying various existential situations people find themselves in. What philosophers can contribute to self-understanding in this process is a clearer view on the questions people may ask themselves when aiming to understand themselves, i.e. the questions generally at stake in existential self-understanding. Wolf, for example, indicates that people who experience meaninglessness or a level of alienation from their lives may ask themselves how enthusiastic they really are about the people and projects they are involved with, or also whether what they are enthusiastic about is worthy of their enthusiasm. People may also ask themselves whether they contribute enough and in a positive enough fashion to the flourishing of what they love. Without providing answers to the specificities of the situations of Anna, Bob, Clemens, Deidre and Edward, Wolf's general criteria may give them direction in their endeavors to make their lives more meaningful.

Frankfurt delineates necessary and sufficient conditions for loving, but these are not as easily used as questions to ask oneself in attempts to love more or in better ways. The main question we can ask ourselves with the help of Frankfurt is: is this love? If our volitional stance towards others is not rigidly focused, then it may be charity rather than love. If our significant others fulfill utilitarian functions in our lives, then at best we do not love them properly and at worst we do not love them at all, according to Frankfurt. He now and again notes how loving may be more ambivalent than the idealized version of loving that he explicates (1999a, 137, 2004, 42, 2006, 44, 51). Time and again these

notes are followed by 'nevertheless' statements though, making it clear that he does not let these phenomena have an impact on his account.

This then is another thing to note about philosophers' contribution to our understanding of existential self-understanding: they proceed by disagreeing with one another. They attempt to figure out on what basis they disagree and argue for the value of their interpretation over another. As said, I think there is much to learn about existential selfhood from Frankfurt's thought. Yet when he writes that love is non-instrumental, I think he describes an ideal, not a reality. I think volition and affect go hand in hand and that 'identification' is too close. Human love relationships are much more ambivalent than Frankfurt's account can do justice to. People question how they should deal with their significant others, which Frankfurt's account cannot elucidate either. Frankfurt's treatment of love is understandable in light of his aim to convince fellow practical philosophers that love is a legitimate source of reasons; in light of the continuity he attempts to ensure between his early and later work through the use of 'identification'; and in light of his main example, which is the love of parents for their newborn and thus far from autonomous children (2004, 43). To some extent, he attempts to do something different from what I intend to do. However, Frankfurt also explicitly aims "to consider the structure and constitution of the self" and argues for his take on it by saying it "pertains more closely [...] to our experience of ourselves and to the problems in our lives that concern us with the greatest urgency" (Frankfurt 1988, viii). For the reasons I just mentioned, I do not think he succeeds at fully doing justice to and clarifying everyday experiences and problems of love and selfhood. Standing on his shoulders, I aim to improve on his analysis in chapter 5.

To allow for the question to arise whether what we love is worthy of our love; to allow for the question to arise how we should relate to our beloved people and projects; to allow for Anna's therapy sessions, Clemens' interview statement and Deidre's ponderings, loving alone is not enough. Beyond loving, people should also be able to form an explicit understanding of the existential aspect of themselves. We now turn to philosophers who have emphasized the structural role self-understanding, taken as a 'narrative' about oneself, might play in existential situations.

2. Being a self by narrating one's self

We are looking at potential contributions by philosophers to our understanding of the existential aspect of the human condition. We have just investigated Wolf's and Frankfurt's thought. Frankfurt aims to elucidate the structural relations between love

and selfhood and does so in a way that (over)emphasizes that it is not up to us to determine who we are. Here I examine narrativist thought, by Marya Schechtman and particularly Charles Taylor.[4] Both of them speak of identity crises and approach (existential) self-understanding as the answer to the question 'who am I?' asked during such identity crises (Schechtman 1990; Taylor 1989, 27). They argue that this answer takes a narrative form, however minimally. Through articulating narratives about themselves do people shape who they are (Schechtman 1996; Taylor 1989). The idea that self-understanding is constitutive of selfhood and that people are thus actively involved in determining who they are is something narrativist theories of selfhood have in common with Søren Kierkegaard's thought. There are important differences too, however, particularly where the role of the social environment is concerned, as we will see in the next section. Here, I will take a brief look at the overarching aims of narrativists before I delve into and evaluate their potential contributions to our understanding of existential selfhood.

Both Schechtman and Taylor claim that current philosophy overlooks (what I would call) the existential aspect to the human condition. Schechtman situates her work first of all in metaphysical debates about personal identity whereas Taylor focuses on practical philosophy. Schechtman (1990) argues against the emphasis within debates about personal identity on the question of reidentification: what makes someone the same person at time points t and t+1? Instead, she maintains, we should inquire into the question of characterization, or self-knowledge: what makes me me (and not you)? Self-narratives give the answer to this latter question.[5] In turn, self-narratives can fulfill the task of what makes personal identity of such practical importance: they can ground moral responsibility, amongst other things. According to Schechtman, they do so in better ways than other accounts of personal identity can (2007). Taylor argues that much of modern practical/moral philosophy overlooks the role of what is of crucial importance to us in our actions, experiences and judgments (1989, 78–87). Additionally, he argues in opposition to the idea that selves could be objects of scientific inquiry (Taylor 1985b, 3–4, 1985a, 1985c). "Human beings are self-interpreting animals" (1985c), Taylor holds.

4 Other notable narrativists are Paul Ricoeur (e.g. Oneself as Another, 1992) and Alasdair MacIntyre (After Virtue, 1984). I focus on Schechtman and Taylor as they provide us with more than enough material for my purposes. MacIntyre's narrativism is closely tied to virtue ethics, which is not my interest here. Ricoeur's work is a universe unto itself; engaging that would make things more complex than seems necessary.

5 Incidentally, Ricoeur argues along almost exactly the same lines for a distinction between idem-identity (at stake in reidentification) and ipse-identity (at stake in characterization). Like Schechtman, he builds his argument in confrontation with the work of Derek Parfit (Ricoeur 1992, 113–39, 1991b).

The stories humans tell about themselves incorporate a sense of what is of fundamental importance to them and thus gives them a sense of direction in their lives.

Conceiving of self-understanding as a narrative emphasizes the active role people play in constructing their selves. Different stories are possible, after all, and the stories we tell about ourselves have consequences. If Bob were to tell himself that his dog has always been a naughty dog, that he has tried over and over again to establish some peace and quiet between them, yet that the dog just would not listen, his current experience of ambivalence is likely to transform into anger at the dog and possibly a sense of relief that the dog is gone. His actions are unlikely to include an extensive search for the dog; he may instead go out to buy a new one. In contrast, Bob may also form a story incorporating his own guilt in treating the dog badly. His experiences may then involve feelings of guilt towards the dog and of shame that he has not been able to take proper care of a dog. He may be inclined to search for the dog to make things right, or he may decide that all dogs are better off without him and spend the rest of his days without one. When we change our narratives, we change ourselves (Taylor 1971, 1985c, 1989).

Schechtman and Taylor emphasize that in this way narratives capture that people actively try to *lead* their lives, beyond simply being alive. Narratives refer to our intentions and goals and the meaning actions have for us in light of them. Whether by reading this you are pursuing your passion for philosophy, doing your job or just passing some time depends on your overall motivations. A mechanical description or even a biological description of actions cannot capture this aspect of the lives of selves: that they experience events in their lives as having meaning in light of goals (Schechtman 2011, 402; Taylor 1971). Goals may be as mundane as 'quenching my thirst' which makes me interpret my colleague's suggestion to get a cup of tea as a welcome break. They may also be of a more existential type.

Taylor's focus lies on goals and motivations that are fundamentally important to people. He refers to what is of utmost importance to us with the term 'the good' and describes what holding an articulated understanding of the good can do. Making explicit what is of crucial importance to us releases its force, he writes, it brings us closer to the good. That is to say, the good's power to motivate us is enhanced by articulation (1989, 92, 96, 107). Deidre is already drawn to philosophy, but explicitly acknowledging this to herself is likely to increase her motivation to pursue philosophy. On the other hand, articulating what is thus important may also lead to inner conflict (1989, 107) as when Deidre reflects on her growing dissatisfaction with her mathematics career, in which she has invested so much. Most of all, it is through forming an explicit understanding of ourselves, of what we love and of how we relate to what we love, that we can evaluate ourselves and our loves. We are not just actively involved in constituting ourselves by forming narratives about ourselves, therefore, but also by evaluating ourselves. This

includes an evaluation of the understanding we hold of ourselves. Not any narrative we form will do and certain narratives are more apt than others. From what we know about Bob and his dog, a narrative in which Bob is at least partially to blame for their bad relationship seems to fit their experiences and behavior better than one in which it is all due to the dog.

Evaluation is never done in a social vacuum. Even when Deidre sits alone in her room pondering how to resolve her professional situation, her thoughts run along lines that have come about in a socio-cultural context. This is a line of thought that runs through all narrativist accounts of self-understanding: just as we learn a language from our social surroundings, we also learn how to narrate from our socio-cultural environment (Schechtman 2011, 402–5) and thus how to understand ourselves. When Taylor further specifies what he means by this, it becomes clear that the goods that he thinks of are values mostly. "People may see their identity as defined partly by some moral or spiritual commitment, say as a Catholic, or an anarchist. Or they may define it in part by the nation or tradition they belong to, as an Armenian, say, or a Québecois. [...] [T]his provides the frame within which they can determine where they stand on questions of what is good, or worthwhile, or admirable, or of value" (1989, 27). Self-understanding according to Taylor is a view of oneself within such a frame of qualitative distinctions. It is a narrative about oneself that evaluates how one is doing with respect to the values one holds due to one's membership of a social group.

Given that self-understanding takes place against a background of socially shared values, and given that self-understanding captures but also constructs selves as subjects of experience and as agents, Taylor writes that people who live in different social environments are in important respects different types of people. "[T]he emotional lives of human beings from different cultures, who have been brought up with very different import vocabularies, differ very greatly" (1985c, 71), he writes. His *Sources of the Self* is in large part an examination of the ways in which people conceived of themselves, i.e. the ways in which they were selves, throughout modern history. When we articulate the self-understanding we implicitly operate on, Taylor says, this enables discussion about what we value. We may first of all discuss this with people who belong to the same social group and share the same frame of qualitative distinctions. We may also talk to people who belong to a different social group, however, and try to convince them of the value of what we value, or discover that they have a better understanding of what is valuable than we do (1971, 13).

When we look at the case of Edward, much of this rings true. In *The Remains of the Day*, the book from which Edward's vignette is drawn, we learn that he was raised by his father who also served as a butler (Ishiguro 1989). Edward thus learned to understand and evaluate himself against a background of values belonging to that

vocation: loyalty towards one's lord, devotion to one's work, dignity. He loves to discuss the meaning of these values with his colleagues. These values are markedly different from the values of wider English society after World War II, where individualism and 'being one's own master' are much more important. Edward indeed leads a very different (emotional) life compared to many of his contemporaries, due to the values he lives by.

Crucially however, Edward's absorption of butler values and his narrative understanding of himself as a good butler have alienated him from himself, particularly from the existential aspect to himself. He has not realized he loves Ms. Kenton, the housekeeper, even though for readers of *The Remains of the Day*, this love is plain and clear given Edward's experiences and behavior in her presence. His narrative self-understanding has captured the values Edward holds, but has not captured what is of crucial importance to him as an individual.

Narrative selves versus existential selves

It seems that narrative self-understanding as envisioned by Charles Taylor captures a socially shared aspect of identity rather than the existential aspect and the way in which this individuates us from one another. As with Harry Frankfurt, the aim to rectify practical philosophy where it neglects the role of what is of utmost importance to people stands in the way of providing a definitive contribution to our understanding of existential self-understanding. Taylor's view on narrative self-understanding conceives of self-understanding against the backdrop of the ethical values our communities hold. What is of existential importance to us as individuals need not feature however, particularly not when we love something or someone that we should not according to the norms and values of our social environment. What is more, the people that we love have much more influence on our self-understanding than other members of our social environment, i.e. those people that we do not really care about. When we evaluate who we are and try to improve our narratives to understand ourselves better, we value the input of the people that matter most to us. Taylor's narrative account does not focus on this either (though see Schechtman 2007, 170).

Narrative views of self-understanding have another limitation when it comes to elucidating existential self-understanding. They guide attention away from the existential content of self-understanding and towards the narrative form. Philosophers debate the question to what extent self-narratives resemble literary narratives (Ricoeur 1991a; Lamarque 2004). Do self-narratives show a measure of thematic unity, possibly by being oriented towards an overall telos (Taylor 1989; MacIntyre 1984; Rudd 2012) or do they not need to do so to count as narratives constitutive of selfhood (Schechtman 2007, 2011)? Also, narrativists generally assume self-narratives to be composed of events in

someone's life history and of the explanatory relations between events (Schechtman 2007), much like the form stories take when people tell about their day ('…and then x happened'). However, to understand *existential* self-understanding, we are not so much looking for stories people tell their colleagues after the weekend, but more for the stories people tell their therapist or their best friends.

In sum, narrativist views of self-understanding aim to elucidate something slightly different from existential self-understanding. Where aims overlap, narrativist views provide relevant insight, e.g. into what articulating can do. Forming an explicit understanding of the existential aspect of ourselves can release the force of what we value, enhancing our motivation to act on its behalf. Becoming aware of what is crucially important to us can also lead to crisis, however, when we wish it were not the case, or find ourselves valuing different things that clash. By explicitly reflecting on what is of existential importance to us, we may resolve issues, tell others about our ultimate motivations and compare and discuss them. None of us lives in a social vacuum. Although our actions and reactions implicitly communicate what we find important, an articulated self-understanding allows for direct interaction with the social environment and the values it holds.

In many cases, aims differ too much, however, for us to gain a detailed understanding of existential self-understanding through studying narrativist accounts of selfhood. Narrativists use their views on self-understanding to make a practical philosophical point and to ground moral agency. Where they do so, they emphasize social values (like 'loyalty' or 'individualism') over other things that may be of value to people (like people they love or pursuits they have invested themselves in). Generally, they emphasize the constitutive role the social environment plays, whereas we see in the case of Edward that what is existentially important to someone need not conform to, or may even be at odds with the values the social environment holds.

What is more, certain aspects of existential self-understanding are not thematized enough in narrative accounts. The role of our relations to what we love does not feature, for example. Also, in emphasizing how we negotiate our understanding of ourselves with others in our social environment, narrativists do not thematize the difference between someone I love deeply versus someone I know somewhat. The former is bound to have more influence on my self-understanding than the latter. And just like there are other things we may find important besides what our social environment finds important, we may also evaluate ourselves along different lines than according to the values provided by our social environment.

3. Becoming a self by relating to oneself

In contrast to narrativists, Søren Kierkegaard and his pseudonyms often warn against the influence of the socio-cultural environment. If you completely interiorize the values of the culture that surrounds you, you may well end up being alienated from yourself. Anti-Climacus, for example, describes a man who, "surrounded by hordes of men, [...] more and more shrewd about the ways of the world—[...] forgets himself [...] does not dare to believe in himself, finds it too hazardous to be himself and far easier and safer to be like the others, to become a copy, a number, a mass man" (1980 [1849], 33–34). Kierkegaard describes self-understanding as something private: it is between you and you (and also, no less private, between you and God).

In brief, being a self is becoming a self, according to Kierkegaard. His pseudonym Anti-Climacus does not speak of self-understanding so much as of self-*relation*. He emphasizes that any understanding we have of ourselves is never neutral; self-understanding incorporates an attitude we hold towards ourselves. This attitude—whether we like what we think we are, or dislike it and try to ignore it, or deny it or try to change it, etcetera—is a major influence on who we are becoming. Whether we are at peace with ourselves or feel alienated from ourselves depends on our self-relation. Self-understanding is not just a cognitive understanding of who you are, it is an affective-volitional-cognitive relation you hold towards yourself.

Anti-Climacus outlines the structure of something akin to the existential aspect to the self and self-understanding. He describes the ideal way in which one can be oneself and, mostly, examines various ways in which people fail to live up to that ideal. In his exposition of various types of self-alienation—or despair, as he calls it—he shows himself to be both a brilliant observer and a stern judge of the intricacies of human existence. His aim and approach are to speak in "the way a physician speaks at the sickbed" (1980 [1849], 5). This is another way in which Kierkegaard distinguishes himself from other philosophers. Just as he emphasizes that self-understanding is not neutral but incorporates an attitude towards oneself that crucially influences who one is becoming, he also does not deem the project of understanding existential selfhood and existential self-understanding to be a neutral academic project. In fact, it should not be, according to Kierkegaard. People suffer from despair and any exposition of why this is so should aim to trigger in readers a reaction that propels them toward a fuller life, more true to themselves.

Kierkegaard thus has both an academic and a therapeutic aim with his work. My own aim to understand existential self-understanding and the situations Anna, Bob, Clemens, Deidre and Edward find themselves in does not have a therapeutic component. Nonetheless, let us look at the way in which Kierkegaard conceives of existential self-

understanding and its impact on the existential aspect of the self. His explications of the relations that constitute us may serve as a relevant addition to the accounts discussed previously. What does his thought have to contribute regarding our understanding of existential situations like those experienced by Anna, Bob, Clemens, Deidre and Edward?

Kierkegaard's pseudonym Anti-Climacus describes self-understanding, and indeed selfhood, as self-relation. To be precise, he describes it as

> "a relation that relates itself to itself and in relating itself to itself relates itself to another" (1980 [1849], 13–14).

This requires unpacking. At a basic level, human beings are *relations*, or syntheses, between opposing poles that Anti-Climacus labels with terms such as 'infinitude and finitude' or 'possibility and necessity'. These syntheses can be out of balance. When people have an abundance of possibility but lack necessity, for example, they may fantasize endlessly about all the things they could become but fail to recognize their limitations and fail to take any concrete steps towards actually becoming anything (1980 [1849], 36). Vice versa, those who lack possibility do not want to imagine themselves and their lives beyond 'what usually happens' and 'how things generally go'. Theirs is a philistine-bourgeois mentality, says Anti-Climacus: they are "completely wrapped up in probability" (1980 [1849], 41). These are the mass men and women that people risk becoming when they only understand themselves in the narrative terms customary to their cultural environment. Both types of people fail to become true selves. They suffer from despair, from alienation of the self, according to Anti-Climacus.

Imbalances do not occur spontaneously. People are themselves responsible for imbalances in their selves through their self-understanding; or, in Anti-Climacus' terms, through the way in which they relate to the synthesis that they are. It is by taking up this *relation to the relation* that people become selves and not just human beings. Anti-Climacus characterizes the self-relation in terms of consciousness and will (1980 [1849], 29). That is, people's selves are characterized by the extent to which they are consciously aware of who they are, but also by the attitude they hold towards themselves that is embedded in their conscious understanding of themselves. When people become aware of who they are, of their limitations and possibilities, they often dislike something they learn about themselves. They may react by trying to forget about the limitations they do not like, or by actively denying their possibilities as those provoke anxiety, etcetera. Their self-relation is thus not a mere cognitive understanding of who they are, but an affective-volitional-cognitive relation.

When we look at Anna's situation, this characterization of self-understanding rings true. Becoming aware of her doubts about the man and the line of work she has devoted her life to makes her feel highly uncomfortable. Her self-understanding clearly incorporates an affective attitude towards herself and not just a neutral cognitive appraisal of who she is. What is more, this attitude shapes her volition and line of action. She wishes the doubts were not there and cannot bear herself very much given that they are there. She tries to deny her doubts at first and then tries to divert her attention away from them. As the doubts do not go away, it is likely that this only adds to her alienation, in line with Anti-Climacus' ideas.

"[E]very moment that a self exists, it is in a process of becoming" writes Anti-Climacus (1980 [1849], 30). People shape the direction in which they are becoming through how they relate to themselves. A crucial part of being a healthy self is being able to accept and deal with the fact that a large part of who we are is not up to us. *We have not established ourselves*, says Anti-Climacus, and there are thus limits to our ability to constitute ourselves. He goes on to infer that *another* must have established us and fills this in in a Christian manner: God has established us. We do not have to accept this further thought, however, when we acknowledge that we have not established ourselves. How we relate ourselves to this fact also shapes who we are. We make take a stance of defiance, for example, and want to determine who we are and are becoming entirely by ourselves. Again, Anti-Climacus says this induces despair and prevents true selfhood. When Anna finally stops denying her doubts and instead faces them head on by going to a therapist, this puts her on a different path of becoming herself, likely a less alienated one.

We have now taken a closer look at all elements of Anti-Climacus' view of the self as self-relation: "a relation that relates itself to itself and in relating itself to itself relates itself to another" who established the relation (1980 [1849], 13–14). The basic relation (or synthesis) that human beings are consists of the (im)balance in our tendencies to dare to imagine what we might be on the one hand and to face what we are on the other. By consciously relating to this synthesis and through the attitude we hold towards it, we establish the synthesis as a healthy or despairing relation. We become selves, in other words, which is the only way to be a self. The self-relation is always also a relation to the fact that we have not established ourselves, i.e. that we are not who we would have chosen to be, had we had the choice. We constantly encounter parts of ourselves that are 'other' in this sense and again our relation to what is other in ourselves constitutes us.

Anti-Climacus speaks of the freedom involved in our ability to relate ourselves to ourselves (1980 [1849], 29). We can influence who we are and are becoming through our self-relation. He emphasizes that this freedom naturally involves responsibility too: "the

self has the task of becoming itself in freedom" (1980 [1849], 35). Anti-Climacus derives this idea of the self as a task from his idea that human beings are intended to be spirit (1980 [1849], 43). That is, human beings are intended to consciously relate themselves to themselves and their lives. Given that we have the ability to not just be alive on auto-pilot but actually *lead* our lives, we should make use of this ability, according to Anti-Climacus. What is more, we suffer from despair when we do not.

If we look at the situation of Bob, we may be inclined to agree with a Kierkegaardian reading of his issues. He does not reflect on his life and his troubled relation to his dog, who is the most important creature in his life. Bob is definitely not a happy man. For that to change, it seems that clearing up his relation to his dog would be vital. In the short run, however, gaining a conscious self-understanding is likely to make Bob unhappier. Bob would become aware of having just one creature in his life that is truly important to him and this creature is a dog. What is more, it is a dog that Bob loathes as much as he loves it. Becoming consciously aware of himself and his life is thus likely to be a painful process. It is hard to see how Bob could pull it off without immediately wanting to forget about all of it again, or forming some other attitude that would lead him further into despair.

Anti-Climacus realizes this. He speaks of different types of despair, all of which come in variations and gradations, depending on the extent to which someone is consciously (cognitively) aware of himself and on the (affective-volitional) attitude he takes towards himself. Anti-Climacus writes that "[v]ery often the person in despair probably has a dim idea of his own state [...] At one moment, he is almost sure that he is in despair; the next moment, his indisposition seems to have some other cause, something outside of himself, and if this were altered, he would not be in despair. Or he may try to keep himself in the dark about his state through diversions [...] through work and busyness as diversionary means, yet in such a way that he does not entirely realize why he is doing it, that it is to keep himself in the dark. [...] There is indeed in all darkness and ignorance a dialectical interplay between knowing and willing, and in comprehending a person one may err by accentuating knowing exclusively or willing exclusively" (1980 [1849], 48).

Some of this can be seen in the butler Edward's case, if not in the brief vignette in the introduction, then definitely in Ishiguro's book on which the vignette is based (1989). Edward has never fully realized that he loves Ms. Kenton, even though his behavior and experiences clearly bear witness to it. At numerous occasions in the book, Edward is 'strangely moved' by the sight of her or by witnessing her emotions, only to quickly move on with his work and forget about it, or rationalize his emotions in professional terms rather than in terms of his feelings for Ms. Kenton. He clearly suppresses any dawning awareness of his love, possibly motivated by his negative ideas

about butlers and housekeepers falling in love, or also by his unease with personal instead of professional relations. When, late in life, Ms. Kenton admits to having hoped to spend her life with him, his heart breaks. Yet before long he is talking pleasantries again, about the need to be grateful for what you have and how one should not think too much about what could have been. Indeed, before long he is by himself again and focusing his mind on thoughts about how to be a better butler. He does not want to know he could have lived (and possibly could still live, if he had the courage) a much happier life than the one he has actually lead. There is indeed 'an interplay between knowing and willing' in his failing self-understanding.

Anti-Climacus does not just describe lived experience, he also judges it and his judgment is strict. Based on his ideas about what counts as despair and what as true, healthy selfhood, he has to conclude that "there is not one single living human being who does not despair a little" (1980 [1849], 22). Had Bob related happily to his dog, had his wife still been alive and had their relationship also been a happy one, and had Bob gone about his days contented and, just as in his current life, without self-reflection, Anti-Climacus still would have judged Bob to be in despair. To be sure, Anti-Climacus acknowledges that "one could humanly be tempted almost to say that in a kind of innocence it [i.e. this minimal version of despair] does not even know that it is despair." However, given that "*veritas est index sui et falsi* (truth is the criterion of itself and of the false)," Bob's situation would be despair nonetheless, according to Anti-Climacus (1980 [1849], 42). Specifically, Bob would be in despair because he does not realize the true meaning of selfhood. He does not realize that he is intended to be spirit, i.e. to relate himself to himself and thus actively take up the task of becoming a self.[6]

Just as lived experience can be deceiving as a basis for deciding whether someone is a healthy self or a self in despair, so too can ordinary language, says Anti-Climacus. He discusses a man who "appropriates what he in his language calls his self, that is, whatever capacities, talents, etc. he may have; all these he appropriates but in an outward-bound direction, toward life, as they say, toward the real, active life" (1980 [1849], 56). This person, Anti-Climacus continues, may be happily married, may be a father, may be someone who is devoted to his work and a well-respected citizen. In other words: he relates himself to his loved ones, he is devoted to his work and uses the capacities and talents he acknowledges are his to these ends. In this manner, he thinks he is a self and he is indeed what his language would call a self. Crucially however, this man does not reflect on himself. He does not take upon himself the task to become

6 Not all of Kierkegaard's pseudonyms are this strict. Johannes Climacus, for instance, excuses 'simple' people from the need to consciously reflect. They may know as much as the wise person, Climacus says, only without knowing that they know (1992 [1846], 160).

himself. He does not realize (whether out of ignorance or because he does not want to or an 'interplay' between these two) that he has the freedom and the responsibility to establish himself as a healthy relation or an unbalanced misrelation. The man lacks inwardness and thereby lacks selfhood, according to Anti-Climacus.

Kierkegaard and his pseudonyms thus employ a different methodological move from the philosophers we have encountered so far. Susan Wolf's ideas about meaningfulness, for example, derive from a procedure of taking *endoxa* as her starting point, i.e. different ways in which people generally speak about meaningfulness in life (2010, 10). She then considers many different examples of life situations that we judge to be meaningful or meaningless. Insofar as the criteria align with our experience and judgments, they stand; insofar as they cannot account for our lived experience and common judgments regarding meaning in life, they need to be adapted. Harry Frankfurt can be said to argue for, assess and adapt his views on love along similar lines. Kierkegaard on the other hand is not persuaded if our experience tells us we are happy and at peace with ourselves. He is also not persuaded by the general ways in which we speak about selfhood. He brings to the table an outsider perspective to judge lived experience and ordinary ways of speaking by: Christianity, or rather, his radical interpretation of Christianity.

Here, Kierkegaard's aim and mine depart from each other. Kierkegaard's ultimate aim is to think through what it means to exist as a Christian and how one can become a Christian (1992 [1846], 17, 249; 1998). He writes that "[i]f people had forgotten what it means to exist religiously, they had probably also forgotten what it means to exist humanly" (1992 [1846], 249) and therefore also closely observes human existential experiences in all their nuances. Yet, Kierkegaard categorizes people's lives as despairing along the lines of his radical interpretation of Christianity, instead of along the lines of lived experience and common ways of speaking about lived experience. (His judgments on selfhood and despair are not even common amongst Christians, as he points out time and again.) I, on the other hand, aim to stay closer to the lived phenomena and ask what different types of inquiry may contribute to our insight into it. Anti-Climacus' relational view of self and self-understanding appears capable of contributing to my project of understanding who we are in terms of what is most important to us. As far as people we love are concerned, for example, we appear to stand in relations to them and Anti-Climacus' views on self-constituting relations may help us understand this better. Again, Anti-Climacus would disagree based on his radical interpretation of what selfhood must amount to. The man who understands himself in terms of his marriage, his fatherhood and his line of work would indeed be what his language calls a self, says Anti-Climacus. Yet none of these facts constitute himself and insofar as he does not consciously relate himself to himself, and thereby does not take upon himself the task to become himself,

he is not a self (1980 [1849], 56). Anti-Climacus actively denies the idea that we could be constituted by our relations to particular people and projects that we love.[7] In this respect, therefore, I can only employ Anti-Climacus' thought to my project against his intentions.

Kierkegaard's outsider perspective on lived experience points to an interesting aspect of the project of trying to understand existential self-understanding. Kierkegaard explicates that what we understand selves to be has an impact on how we can be selves. So not just our understanding of *ourselves*, but also the project of understanding *the (existential) self* influence who we are. If people understand healthy selfhood to be selfhood judged positively in the eyes of their social surroundings, they will be different from when they understand healthy selfhood to be selfhood judged positively in the eyes of God (1980 [1849], 79). Put differently, if people think that in order to understand selfhood, one should inquire into what people generally think selfhood is, they will be selves in different ways from when they think that in order to understand selfhood, one should take to heart Christian scripture.

A natural science of existential selfhood

In his notebooks of 1846, Kierkegaard writes of another possible source of insight into the nature of the self: natural science (2011, NB 70 - NB 90). The authors of the textbook in cognitive neuroscience quote from these notebooks, as we saw in the introduction (Gazzaniga, Ivry, and Mangun 2002). Although the discovery of functional imaging techniques relying on blood flow was a long way off still, Kierkegaard imagined an instrument that would allow scientists to 'hear the brain beat', much like the stethoscope allowed people in his days to hear their heart beat (2011, NB 76). He envisions a 'physiologist who counts pulse (beatings) and studies nerves' (2011, NB 70) and imagines future physiologists who may make it their task to explain what love is or what freedom is (2011, NB 70). Kierkegaard is wary of people thinking that such a natural science of spirit (2011, NB 78) would have contributions to make to our understanding of existential self-understanding. He argues, first, that science distracts from the task to be an existential self. Natural science "captivates, interests, elevates, enchants, teaches, enriches, seduces, convinces" (2011, NB 83). It takes up time and takes your attention away from what matters most: dealing with yourself and your life (2011, NB 78).

What is more, just like socio-culturally common ideas about selfhood foster an attitude of looking at what others are doing to figure out how to be oneself, natural

7 For a full argument that Anti-Climacus denies that true selfhood can be constituted by our relations to beloved people and projects, see my 'Selves, existentially speaking' (van Stee 2015b).

science also fosters a wrong attitude, opposed to true selfhood, according to Kierkegaard. Science promotes a disinterested attitude; it promotes neutrality, objectivity, skepticism and a postponing of judgment until further evidence comes in. Whereas for tackling existential questions you need to be able to decide and act in the face of uncertainty. You need enthusiasm and courage to be able to do so. That is, you need to be in the opposite state of mind from the cold neutrality and postponing of judgment that science requires and encourages (2011, NB 70, 73, 79).

Furthermore, Kierkegaard posits that science introduces a distinction between what he calls 'the simple' on the one hand and 'the learned and half-studied' on the other (2011, NB 84). All people are equally (in)capable of dealing with themselves and their lives, Kierkegaard says. Science, however, differentiates between people, between the simple people on the one hand and, on the other hand, those who dedicate their lives to learning or have studied somewhat and looked through a microscope once. The authority of science may make it seem as if we had better listen to the learned experts when needing to make existential decisions, or as if those who have 'looked through microscopes' are better capable of existing than those who never studied. Kierkegaard ridicules this idea and finds it dangerous.

When discovering her dissatisfaction with mathematics and her attraction to philosophy, Deidre could have read up on average levels of happiness amongst professional philosophers and mathematicians. She could have postponed her judgment until she would have been absolutely certain that quitting mathematics and studying philosophy would lead to a fulfilling and useful life whereas continuing her work as a mathematician would lead to utter dissatisfaction. Who knows, the thought could have crossed her mind that some neuroscientist should scan her brain, to see whether areas associated with love and happiness light up more when she is working on mathematical problems or philosophical analyses. According to Kierkegaard, how we think we should go about gaining understanding about ourselves matters for self-understanding. If we think we should defer to scientific knowledge, or to the opinion of society, our self-relation will be different from when we think the only one who truly knows us is God and, as Anti-Climacus would have it, self-understanding can best be gained by relating ourselves to ourselves before God.

This is also why Kierkegaard builds up his authorship in the way that he does. His pseudonyms are different personas who write from somewhat different perspectives. Anti-Climacus speaks as if he has grasped it all and considers himself a truly Christian human being. Johannes Climacus is more modest: he writes as he has cognitively understood something that he deems too important not to share, yet emphasizes that he is unable to live according to the radical standards he sees in Christianity (1992 [1846], 617; 1980, xxii). Kierkegaard communicates indirectly, not just by using pseudonyms, but

also through the use of irony and through literary means. In doing so, Kierkegaard tries to attune people to what he deems to be true selfhood. That is, he tries to get his readers into the right state of mind to take the task of becoming themselves upon themselves.

In sum, Kierkegaard and his pseudonyms understand selfhood in terms of self-understanding as self-relation. They understand self-relation not along the lines of how people generally understand themselves and relate to themselves, but in terms of how people *should* understand themselves and relate to themselves. Kierkegaard's aims thus do not coincide with mine. I wish to elucidate the existential aspect of the human condition *as it is lived* by people, instead of judging Anna, Bob, Clemens, Deidre or Edward, as Kierkegaard would. Kierkegaard does not just articulate an ideal, however, he also closely observes various ways in which people actually relate to themselves. This allows his thought to elucidate existential selfhood in several ways; I will build on these ideas in chapter 5.

First of all, his relational view of self-understanding is interesting as it emphasizes that self-understanding is not just a cognitive representation of selfhood, but also an attitude we hold towards ourselves. Kierkegaard emphasizes this more strongly than narrativists like Charles Taylor do and thinks through the consequences the attitudes have for people's mental health. For Kierkegaard, these self-relations are what constitute selfhood. Whereas Harry Frankfurt conceived of self-constituting relations in terms of volition, Kierkegaard emphasizes how the attitudes we hold towards ourselves are affective-volitional. Although he does not say of relations to beloved people and pursuits that they truly constitute selfhood, I can make use of his insight on self-constituting relations against his intentions on this point.

Kierkegaard does not develop his views in terms of necessary and sufficient conditions, as Wolf and Frankfurt do. Instead, his view on the structure of selfhood and self-understanding accounts for the dimensions along which (despairing) versions of self-relation differ and how one version can develop into another. We can elicit questions from his views that people who try to understand themselves can ask themselves. Given the close link of Kierkegaard's views with issues of mental instability, such questions may also be relevant to therapists treating clients with existential issues. 'What can you imagine becoming, say if your life were to change drastically?' and 'What would you always be, even if your life were to change drastically?' are questions pertaining to what Anti-Climacus calls the synthesis human beings are. 'Do you feel you have had an influence over who you have become so far?' and 'Do you feel responsible for who you will become in the future?' are questions asking after the extent to which people experience, in Anti-Climacus' terms, the task of becoming themselves. People can be asked to reflect on parts of themselves that they would not have chosen, had they

had the choice, or on experiences that shaped them that were outside of their control; these questions pertain to what is 'other' in people's selves. Naturally, they could also be asked to reflect on what they like about who they are and what they do not like, and why, corresponding to the self-constituting attitudes people hold towards themselves. All these questions can be relevant in people's attempts to gain an understanding of the existential aspect of themselves, irrespective of whether they accept the substantial, radically Christian direction of Kierkegaard's therapeutic project.

Furthermore, Kierkegaard's views are relevant for the way in which they think through the becoming nature of the self. Being a self means becoming a self, Anti-Climacus emphasizes. Self is thus a process, not a substance. Kierkegaard and his pseudonyms develop an account of a rich, full-blown existential self without falling into dualism. It is the way in which people relate themselves to themselves that determines the direction in which they are becoming.

Which brings us to the final point of insight Kierkegaard's thought can contribute to understanding existential self-understanding. He emphasizes the practical import, not just of self-understanding, but also of understanding self-understanding. It matters whether we articulate understandings of the existential aspect of the human condition in terms that are close to human experience and recognizable because of it; or in terms that belong to the technical vocabulary of a natural science and are thus only truly accessible to experts; or in terms that articulate an almost unattainable ideal in light of which we cannot help but falter.

4. Philosophy's contributions to understanding existential selfhood

Human beings tie up their identity with people and projects that they care about deeply. When they reflect on who they are, they often understand their identity in terms of what is of utmost importance to them. This may give them a sense of direction in life, a basis on which to make important decisions. Irrespective of conscious awareness, people are naturally motivated to act to the benefit of what they love. Their experiences too are affected by the lot of their beloved people and projects. I aim to understand this existential aspect of the human condition and the types of insight into it that cognitive neuroscience and philosophy have to offer. Anna, Bob, Clemens, Deidre and Edward provide the phenomena that serve as a touchstone throughout.

In the previous chapters we saw how cognitive neuroscience (CNS) mainly delivers insight into the neural underpinnings of love and self-reflection. Interpretation plays a crucial yet largely unacknowledged role in its research process. Both the lack of

critical reflection on conceptual assumptions and incentives towards standardization of interpretations prevent the extension and improvement of CNS research. So far, CNS appears to have little to offer in terms of improving our understanding of love and self-reflection, or of human beings as loving and self-reflecting beings. It only adds knowledge (or, at this stage, hypotheses) about the neural networks that are active during love and self-reflection.

In this chapter, we looked at several philosophers who think through aspects of something akin to existential selfhood and existential self-understanding. All of them have some insight to offer. Susan Wolf unearths conditions on which our judgments about meaning or meaninglessness in life are based. This helps to understand those judgments better. It also provides us with questions people may ask themselves when they experience meaninglessness or otherwise aim to make their lives more meaningful. The necessary and sufficient conditions for meaning are not just conceptual but also describe a view of the good life. Harry Frankfurt thinks through the structural relation between love and selfhood. He outlines necessary and sufficient conditions for loving in light of its structural connections to our practical identities. In the process, he shows how different experiences people live through are connected to each other, even certain experiences that appear to contradict each other (e.g. love captivates us, love sets us free). Charles Taylor reflects on how self-understanding can give direction in life. He examines what having an explicitly articulated self-understanding makes possible. He conceives of self-understanding as a self-narrative composed within a framework of overarching values that we hold due to belonging to a socio-cultural community. Søren Kierkegaard, finally, describes how self-understanding always involves a relation to ourselves that influences who we are and are becoming. He describes the various ways in which we may not be at peace with ourselves and makes us understand how they are structurally connected to each other.

These philosophers have objectives with their views that are different from mine. When evaluated in the context of my project (and sometimes also in the context of their own projects), their views have limitations. Wolf does not provide an analysis of the structural relations between love and selfhood. Frankfurt does, but in his view we identify ourselves with our beloveds and their interests, and cannot help but accept their interests as our own. He thus takes away the distance between ourselves and our beloveds that allows for a measure of independence from them. His views cannot account for ambiguities in love relationships and the way these influence people's experiences and actions towards their loved ones. What is more, Frankfurt's claim that loving is volitional first and foremost makes experiences of alienation and of being at peace with oneself harder to understand. Insofar as Frankfurt aims to elucidate experiences and problems that are part of everyday human life, his account does not live

up to his own aims. Taylor focuses on the way in which the values that we hold shape our practical identities. Whereas what we love tends to individuate us, values form a socially shared aspect of our identities. Kierkegaard, finally, does not aim to elucidate lived experience as it is experienced and talked about by those involved. Instead, he aims to judge lived experience from the perspective of a substantive view of what selfhood should be like, which is a radically different perspective from the one generally taken up when trying to understand existential experience.

Generally speaking then, philosophical inquiry into existential selfhood offers several types of contributions to our understanding of existential self-understanding. First, philosophers describe lived experience in its structural characteristics. They thus deepen our understanding of ourselves, our experiences and actions. They point out phenomena that we live through but glance over in daily life. Second, philosophers analyze the structural connections between lived experiences. They articulate experiential structures present in and forming the condition of possibility for different types of lived experiences (Zahavi 2005, 97). I already mentioned how Frankfurt connects different phenomena, even apparently contradicting ones, such as experiences of love as both captivating and liberating. When Kierkegaard gives a description of the structure of selfhood as a relation that relates itself to itself (etc.), and of the self-relation as constituted by an interplay of knowing and willing, this enables him to place all sorts of despair in relation to each other, depending on whether they are characterized by a lack of knowledge about selfhood, a lack of knowledge of oneself, a lack of will to be oneself, or a will to be oneself but inability to accept that which is other in oneself, and various interactions between these.

Third, philosophers unearth the assumptions underlying common ways of thinking about lived experiences. Wolf does so in her analysis of why we judge certain lives to be particularly meaningful and others not so much. I did this in the previous chapters when I unearthed conceptual assumptions embedded in operationalizations of love and self-reflection. Fourth, philosophy is also a critical enterprise: philosophers may question the assumptions they unearth or argue against a particular way of employing a concept. In the first part of this chapter, I criticize Frankfurt's analysis of loving and of the structural relationship between love and selfhood, for example. And in the previous chapters I pointed out the limitations of current operationalizations of love and self-reflection. Finally, philosophers point out, analyze and develop views of the good life. In other words, philosophers develop normative views on human existence. Kierkegaard is a clear example of a philosopher who proclaims a view of the good life and, on that basis, judges common human behavior and experiences to be deficient. Yet when Wolf unearths assumptions underlying common judgments about meaningful lives, she is in effect also unearthing a view of the good, meaningful life.

Clearly, there is more to be known about existential selfhood than what philosophers are able to provide. For one, philosophy does not take into account the particularities of different people's lives. Anna, Bob, Clemens, Deidre and Edward need to figure out how to deal with their situations. Philosophy is not going to provide clear-cut answers to their questions. Instead, philosophy may clear up the ways in which they can think about their predicament and point out some of the questions they may ask themselves to help them decide how to move forward.

What is more, philosophy does not provide insight into the biological enabling conditions allowing people to experience love, or meaning in life or an identity crisis. Philosophy does not provide insight into the brain activity that correlates with self-reflection or love either. That is to say, philosophers of existential selfhood are interested in fundamentally different questions than the ones addressed by CNS research. CNS and philosophy are thus not in competition with each other, as some of the authors that we encountered in the introduction would have it (Churchland 2008; Farah and Heberlein 2007; Gazzaniga, Ivry, and Mangun 2002) but address different questions. In chapter 4 I address the future potential of CNS to contribute to self-understanding.

5. Interpretation in philosophy and cognitive neuroscience

In all their research, philosophers rely heavily on interpretation. Charles Taylor writes how "interpretation [...] is an attempt to make clear [... that] which in some way is unclear. The interpretation aims to bring to light some underlying coherence or sense" (1971, 3). He may as well have given a broad description of what philosophers do; it sounds similar to my description of existential philosophy's aim as the uncovering of underlying structural characteristics of existential phenomena, thereby elucidating those phenomena. Taylor emphasizes that any interpretation is by a subject (1971, 4–5). That subject has certain words at her disposal, but not others, and these have certain meanings, but not others, due to the language environment she belongs to. What is more, several interpretations of underlying coherence or structure of a phenomenon may be possible and the aims philosophers have in mind influence the interpretations they put forth. This is why in trying to understand a particular philosophical view on existential experience, it is important to pay heed to the overarching aims the philosopher has with her work.

In CNS, interpretation plays a role at two different moments in the research process and it takes a different form. First, researchers need to concretize their topic of interest into a task, operationalizing love or self-reflection. Any task embodies only one

of several possible interpretations of that phenomenon. Indeed, it often embeds only one of several aspects of the phenomenon. In this sense, operationalization involves an interpretation of the phenomenon of interest. Unlike philosophy's interpretations, CNS tasks are not meant to clarify the phenomenon, but are meant to be used. The aims researchers have in mind when choosing one possible operationalization over another may have to do with their ideas about love or self-reflection, but more often than not they come down to issues of usability. Second, researchers draw informal reverse inferences from neural data to ideas about love or self-reflection. In effect, they interpret the phenomenon of interest in light of the new, neural information they have found about it. As we saw in chapters 1 and 2, information about neural correlates currently cannot be translated reliably into insight regarding love or self-reflection. In choosing one interpretation over the other, researchers rely on what others have done and on preconceived notions of what they deem love and self-reflection to be like.

Even though interpretation of love is somewhat different in philosophy of love and the two moments of interpretation in CNS research on love, it always has conceptual consequences, highlighting a particular angle on love, but not others. What is more, unhelpful strategies to deal with interpretation are the same for both disciplines, as are preferable ways of tackling interpretation. When it comes to interpretation, there is no standard set of steps to follow, no methodology in the way there is for part of the research process in experimental research. Rather, the proof of the pudding is in the eating. A particular interpretation is recognized to be apt or not or only in some limited way. Once a particular interpretation is on the table, critical reflection and discussion about it are possible. It is the existence and quality of those discussions that further interpretative work.

Cognitive neuroscientists generally do not explicate which aspect of love their tasks focus on or how that aspect relates to others. They mostly do not argue why that aspect is focused on and not others. Instead of critically examining their colleagues' interpretative choices in task development and discussion of results, they often copy them. As said before, cognitive neuroscientists are not trained to deal with their moments of interpretation critically; they are mostly unaware of the impact of these moments in the research process. For good reasons, much of their training focuses on mastery of the imaging techniques and data analysis practices. On those points, standardizing research practices makes a lot of sense. Where it comes to moments of interpretation, however, critical reflection is key.

Philosophers only start their work when a first description of a phenomenon, a first explication of a concept, or a first normative view on the good life is on the table. They try to articulate this view as clearly as possible. Philosophers ask themselves: what do I mean with my terms when I say this? They try to rid their views of flaws by

systematically questioning them. What am I presupposing when I say this? And is what I am presupposing internally coherent? Can I think of phenomena that disprove my interpretation and if so, how do I have to adapt my interpretation such that it can accommodate them? How does my view fare in comparison to other views on this phenomenon, both historical and contemporary, and what are my arguments for preferring my view over others?

To some extent, the self-critical move takes place behind the scenes, before a philosopher writes up her views. Part of it is on display in the argument she puts forth. Critical reflection also takes place after she has presented her views, with the help of colleagues. Colleagues may be better able to see the weaknesses in an argument, to come up with counterexamples and to present alternative views in the strongest possible light. Philosophers thus proceed by disagreeing with each other. Ideally speaking, they analyze the basis on which they disagree and deepen their understanding through that. They come up with different possible interpretations of a phenomenon and then try to argue for their interpretation or to uncover under what circumstances which interpretation is most apt. Philosophers interested in the human condition generally, or in the existential aspect of the human condition specifically, constantly refer to lived experience in their research process. The justificatory strength of their analyses lies in the extent to which they clarify lived experience and deepen our understanding of it. That is to say, they test their ideas against experience.

It is crucial therefore, that these philosophers are aware of variations in human experience and capable of reflecting on them. That the system is not watertight on this point can be glanced from history. Historically, philosophers were overwhelmingly male and white and many of them failed to see that women or people of color were as fully human as themselves. Many discredited important aspects to the human condition that were associated with what they were not, like emotionality, and forgot to think through important human issues that did not feature prominently in their own lives, such as parenthood. Exceptions notwithstanding, the study of human existence was impoverished through the homogeneity of its practitioners and, just as importantly, the fact that these philosophers did not listen enough to people who lived different human existences from their own. Although current philosophy is less homogeneous than it used to be, biases and blind spots remain a thing to watch out for. Like cognitive neuroscientists and their research participants, philosophers are generally WEIRD, i.e. they are from Western, Educated, Industrialized, Rich, Democratic societies (Henrich, Heine, and Norenzayan 2010). What is more, inclusion into academia requires high levels of productivity, a willingness to live through a relatively long period of financial insecurity and an ability to appear to those already working in academia as someone who fits in, i.e. as someone who shares certain social mores. These factors may make it

hard for those from lower class and lower middle class backgrounds to enter academic philosophy, for example, as well as for those who are unlucky in terms of health or the burdens they have to carry next to their academic work.

However, philosophical analyses of the human condition need not suffer necessarily if the people working as academic philosophers are able to empathize enough with others to put their issues on the agenda and to think through them fully. A philosophy education is partly that: a confrontation with many different ways of thinking about the human condition and the questions it poses. What is more, philosophy's research practice encourages self-critical reflection. It encourages philosophers to come up with and discuss strong counterexamples to their own view. It encourages colleagues to do so. In principle, philosophical research practice encourages looking at things from a different perspective and reflecting on how different perspectives may relate to each other. Although practice may not always follow principle, this is a big difference with many empirical sciences, including cognitive neuroscience. As we saw in the first chapters and will see in the following chapter, current research practice in CNS encourages copying of interpretations instead of critically reflecting on them and diversifying them.

All in all then, the idea that philosophy is a speculative prelude to experimental research, gradually "ced[ing] intellectual space to increasingly well-grounded experimental disciplines" (Churchland 2008, 409) is misguided. We saw before how philosophy and CNS address different questions. We see here how interpretation plays a role in both research practices. Philosophy depends heavily on interpretation and has developed strategies of critical reflection, collegial dissent and diversification of interpretations to ensure the quality of its analyses. The quality of cognitive neuroscience research also depends on the quality of the interpretations researchers inevitably make, e.g. in developing tasks. Cognitive neuroscience so far has not developed any strategies to ensure the quality of interpretations and its research process contains counterproductive incentives on this point.

Yet of course cognitive neuroscience has only been around for twenty years (and then some), whereas philosophy has been around for two millennia (and then some). CNS may not greatly contribute to self-understanding and our understanding of self-understanding yet, but it holds great promise for the future. Or so many argue. In the next chapter, I examine what CNS may contribute to our understanding of self-understanding in the future, particularly if it were to take inspiration from philosophy in the way it deals with its moments of interpretation.

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 embrained 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 embrained 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 embrained 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?

5

Understanding Existential Self-Understanding

Human beings are often not entirely at peace with themselves. Deidre, for example, experiences restlessness concerning her career in mathematics and her attraction to philosophy. Sometimes people experience profound dissatisfaction, as Bob does, who spends his days in frustration and anger. They may even experience outright alienation. Anna wonders what it all is for: why go to work every day? Why stay with her husband of many years? In these cases, as well as in the cases of Clemens and Edward, the dissatisfaction has to do with people they have tied up their identity with and pursuits they have invested themselves in. That is to say, their existential misgivings have to do with the people and pursuits that are of utmost importance to them. Maybe they simply experience frustration, as Bob does. Yet they may also reflect on their situation and wonder what to do about it. Should Anna leave? Should she somehow change perspective on her life?

I aim to understand existential self-understanding. Existential self-understanding is an aspect of the human condition that is at play in the situations Anna and the others are in. In this study, I approach 'existential selfhood' and 'existential self-understanding' in a bottom-up fashion, if you will. Instead of defining them at the outset, I investigate the contributions cognitive neuroscience and philosophy have to make to our understanding of existential self-understanding. In the previous chapters, we have seen how different the contributions are that both academic disciplines can make. Cognitive neuroscience (CNS) investigates the neural activity that correlates with behavior and experiences. It thereby provides insight into the enabling conditions underlying lived experience, but not into the reasons why we behave as we do, as explained in the previous chapter. The philosophers discussed in chapter 3 aim to elucidate lived experience. They analyze structural characteristics of various existential experiences, behaviors and problems. They unearth assumptions underlying ways of talking about the existential aspect of the human condition and critically scrutinize such assumptions too. They also analyze and develop views of the good life, of how we should deal with the existential issues life confronts us with. Yet the philosophers that look into the existential aspect of selfhood have done so with aims in mind that are different from mine. Often, their normative ambitions get in the way of getting the phenomena in full

view. As a result, they do not elucidate the existential phenomena that are key here, or do so only partially.

In this chapter, I draw lessons from the previous chapters regarding the existential aspect of the human condition, in light of the cases of Anna, Bob, Clemens, Deidre and Edward. I develop a philosophical anthropological view of existential selfhood and existential self-understanding and draw on empirical research where that seems appropriate. Mostly, I synthesize the different strands of thought on existential selfhood that I discussed in chapter 3 and try to improve on them and add to them where necessary. One recurrent question is in what ways the shaping of existential selfhood happens to us and how we are nevertheless implicated in it too. Harry Frankfurt's ideas about love emphasize that we cannot help but be who we are, existentially speaking. Charles Taylor and particularly Søren Kierkegaard's pseudonym Anti-Climacus accentuate that we are actively involved in the constitution of ourselves. From the perspective of the first approach, the idea that we constitute ourselves may make the self seem too fickle, i.e. too changeable to form the basis of our identities. From the perspective of the second approach, on the other hand, the worry rises that if self-constitution were not up to us, this would surrender the self to the people and pursuits it is attached to, not doing justice to the measure of independence that is implied in the notion of selfhood. When we look at Anna, Bob, Clemens, Deidre and Edward, we clearly see that they have to deal with an existential situation that is in important ways not of their doing, yet that they still can do something about.

I shall argue that existential selfhood is formed by the relations in which people stand to what is of fundamental, personal importance to them. These relations are characterized by the affective-volitional stance people hold towards what is of fundamental importance to them, as well as by the cognitive understanding they hold of these relations. The affective-volitional stance (as witnessed in people's actions and emotions) and the understanding people hold of it form a constant loop, such that being oneself is always a matter of becoming oneself.

I unpack these ideas one by one. First I explicate what I mean by 'fundamental personal importance'. Then I address 'relations' and their affective-volitional quality. For both, I make clear how I draw on Frankfurt's work, yet deviate from his views too. Subsequently, I discuss the cognitive dimension that is added to the relations through self-reflection and the constant loop it forms with the affective-volitional relations it forms an understanding of. Both Taylor and Kierkegaard are my discussion partners here. Finally, I draw conclusions. By explicating structural characteristics of the existential aspect of the human condition, I in effect elucidate the questions that are structurally at stake in existential situations. Bringing these questions into clearer view is valuable in its own right, as it clarifies the questions we (including Anna and the

others) may ask ourselves when faced with existential situations that need resolving. It is also relevant as preparatory work for the development of normative views on how we should lead good human existences, and optionally, as preparatory work for front-loading phenomenology into CNS research design.

1. 'Fundamental, personal importance'

Bob is attached to his dog, but his behavior towards the dog is highly ambivalent. The people who are of utmost importance to Clemens have died. Edward allows his professional commitment to overrule any other interest he might have. Bob's dog, Clemens' wife and children, Edward's profession as a butler: these are the people and pursuits that are fundamentally, personally significant to Bob, Clemens and Edward. This shows itself in their actions and experiences, including the problems they experience. Most of Edward's daily *actions* are in accordance with the demands of his profession as a butler. The frustration and anger that Bob *experiences* throughout the day are tied up entirely with his dog. Clemens' identity *crisis* would not have arisen, were it not for the fact that the people who are most important to him are dead. In general, what is existentially important to people provides them with important motivations for action. The fortunes and misfortunes of important people and projects do not leave them cold: they cannot help but be emotionally affected by them. What is more, that these people and projects matter so much also means that when relations to them are for some reason troubled, existential unease is bound to arise. Vice versa, having important people and projects in their lives and having good relations with them, contributes to giving people a sense of meaningfulness.

Existential importance is inherently personal. Food and water are important too of course. When they lack, we are motivated to pursue them, and until we find them, all our experiences may be characterized by a type of cranky restlessness. Yet, where food and water are concerned, this is the same for all of us. In contrast, what is of existential importance to us is different from person to person. It individuates us into different selves. Clemens' identity is thoroughly shaken by losing his wife and children, and he grieves them intensely. The chef of the supermarket at which the family regularly shopped is unlikely to experience an identity crisis after their deaths, however. She may feel very sorry for the family, but to her, they were likely just members of a large group of customers. Personal importance is 'rigidly focused', to use Harry Frankfurt's terminology. It concerns this particular person, not any person of a certain class, say of customers. And people focus rigidly on different people and different pursuits. In this

sense, the set of people and pursuits that are existentially significant to them individuates them.

Existential importance is a gradual notion. Several people and pursuits can be important to us, yet there are bound to be differences as to how important they are to us exactly. Even if we are affected by the fortunes and misfortunes of various people, we may find that the fate of person A affects us deeply, whereas upon hearing what happened to person B, we quickly move on. Likewise, we are immediately inclined to help person A, even though we were planning to spend that time on a project that is also important to us. Whereas when it comes to person B, we prioritize our project instead, hoping that somebody else will help B out. Importance is a gradual notion and it may not be possible to come up with a clear cut-off point.

Naturally, there are differences between fundamentally important people and fundamentally important pursuits. Developmentally speaking, crucially important people are present from the beginning of our lives, whereas projects are not. Also, people talk back, act back, have their own experiences. They are agents themselves, and are thus actively involved in the relation in which we stand to them. Projects, professions and other pursuits that are utterly important to us, on the other hand, depend on our agency for their existence.

It is to a large extent not up to us to decide who and what is thus important to us. Here we thus confront a structural characteristic of existential selfhood that is not ours to establish. We discover what is crucially important to us, more than we decide it. Of course, we can seek out circumstances that make it likely that we start finding someone or something important: we can sign up for a dating site, get a dog, sign up for an intensive ceramics course and day dream about owning a little shop where we sell personally designed, handmade earthenware. Yet, it may turn out that we do not much take to the dog. It may turn out that we find turntables much harder to handle than we had anticipated, and pottery frightfully boring. We can never determine at will to be attracted to, let alone fall in love with a particular individual. Similarly, we cannot stop finding particular people and projects important at will. The marriage may be over, but that does not mean that we are no longer attached to our former partner. For better or worse, we are still affected by how they are doing. Certain people are with us for life in this respect, it seems, even when they die. We remain attached and our relation to them remains constitutive for who we are. In short, we cannot decide at will who and what is of such importance that our existential identities are shaped by them.

Up until here, Harry Frankfurt and I largely agree; I draw on his work. Frankfurt speaks of 'love' though, whereas I prefer to steer clear of that term, using 'fundamental, personal importance' instead. Frankfurt thinks through an idealized version of love, for example

when he emphasizes its non-instrumentality. As I noted in chapter 3, Frankfurt has two aims with his work that are sometimes in tension with one another. On the one hand, he wishes to convince practical philosophers of the relevance of 'what we love' to theories of practical reasoning (1999b, x). In all likelihood, they are more easily convinced that love can provide us with legitimate reasons for action if love is treated as an ideal and if love is more closely associated with stable volitional tendencies than with loving feelings that may prove unreliable. On the other hand, Frankfurt aims "to consider the structure and constitution of the self" in such a way that it "pertains [...] to our experience of ourselves and to the problems that concern us with the greatest urgency" (1988, viii). This aim, of a more phenomenological than normative bend, if you will, aligns with mine. I aim to elucidate the structural characteristics present in existential experiences, actions and problems. Idealized notions of what love should be like do not serve my purposes, quite the contrary. I choose not to use the term 'love' as it carries so many positive connotations. Self-constituting relations to important people may not be all that positive, however.

2. 'Relations': affective-volitional

Existential selfhood is formed by the relations in which people stand to what is of fundamental importance to them. When people or pursuits are important to us, a relation between us and them is in place: we are affected by their fortunes and are motivated to act so as to influence their lot. The actual character of our emotions and actions with respect to them depends on the character of the relation, however. I maintain that relations are affective-volitional entities. That is to say, they are characterized both by the feelings someone holds to what is important to her as well as by what she is motivated to do with regard to the other. In yet other words, it involves the particularities of her emotional life as well as of her practical identity.

Affect and volition are inextricably entwined. Bob's relationship with his dog is troubled. Bob's existential identity is constituted by his relation to his dog, and is thereby equally troubled as that relation, especially given that Bob has no other important beings or projects in his life. His emotional life is a whirlwind: when the dog is still there, he is often angry at it; when the dog disappears, he is crying in his room, utterly distraught. Bob's practical identity is characterized by ambivalence: he walks the dog devotedly, but kicks the dog and yells at him during these walks. When the dog has disappeared, he does not know whether to go look for him or not. It is exactly at those moments that Bob is infuriated with the dog (affect) that he wishes him ill and kicks him (volition and

its practical consequences). When he feels sad because he misses the dog (affect), he vouches to search for it after all (volition). Structurally speaking, affect and volition are intertwined in the way they constitute the existential aspect of the human condition.

Although I build on Frankfurt's important work, here I clearly depart from him. As I explained in chapter 3, Frankfurt conceptualizes the relation between love and selfhood in terms of identification. When he describes how "a lover *identifies himself* with what he loves" he states that "[t]he interests of his beloved are not actually *other* than his at all. They are his interests too." And a bit later: "its interests are identical with his own" (2004, 61–62, his italics). Furthermore, Frankfurt treats loving as a volitional matter first and foremost, writing that it consists in a practical concern "for the existence and the good of what is loved" (2006, 40). He also writes that "enthusiasms are not essential. Nor is it essential that a person likes what he loves. [...] the heart of the matter is neither affective nor cognitive. It is volitional" (2004, 42, 1999a, 129, 1999c, 161). I, on the other hand, think that existential selfhood is formed by the *relations* in which people stand to significant others instead of by identification with them and their interests. What is more, I think these relations are best characterized as *affective-volitional*.

Let us take another look at Bob and his dog. The dog is the most important creature in Bob's life. Yet Bob does not include the dog's interests amongst his own, or at least not unambiguously so. The dog's wellbeing may be supported by Bob walking his dog, it is most definitely also hurt by him kicking his dog. There appears to be more distance between Bob and the dog, particularly between Bob's interests and the dog's interests, than Frankfurt's 'identification' can account for. 'Relations' is supposed to capture this distance. It provides the locus where various types of ambiguity and the resulting modifications in existential selfhood take hold. If Bob stands in a love-hate relationship to his dog, then his affective experiences regarding the dog are bound to involve love and hate. Bob's actions are clearly characterized by ambivalence too. His selfhood is formed by the love-hate for the dog, not by an identification with the dog's interests. (For alternative readings of Bob's situation and rebuttals, see Van Stee (2015a).)

To put it even more extremely, also people we absolutely hate are important to us. Their fate does not leave us cold, albeit that in such a case we may rejoice when they fail. We are automatically motivated to act in order to have an impact on them too, albeit to their detriment. That is to say, ambivalent relations and straight-out negative relations to people also form the existential part of ourselves. Frankfurt's identification account cannot accommodate such formative influences on our selfhood. A view focusing on relationality can. The examples also demonstrate that we need to take both affect and volition into account when aiming to describe the structural ways in which our selves are shaped. Enthusiasms for important people and pursuits may not be essential for them to have a formative influence on who we are. Yet depending on

whether we are enthusiastic about them or not, the character of that influence is bound to be different.

Hatred and love-hate may be rather extreme examples, but everyday loving relations are generally more ambivalent than the idealized version of love that Frankfurt describes. If we want to elucidate how everyday relationships influence who we are, there needs to be a locus for ambiguities to take hold. What is more, I would argue that even perfectly pure loving versions of relationships and their effects on our identities are better understood through an affective-volitional account rather than a volitional account. After all, Frankfurt writes that a lover thrives or suffers depending on whether his beloved thrives or suffers. It is hard to understand this impact on the emotional wellbeing of the lover in volitional terms only.

Finally, a view in terms of affective-volitional relations is better able than an identification view to elucidate the main phenomenon at stake here: the inchoate dissatisfaction with and outright alienation from oneself and one's life that Anna, Bob, Clemens, Deidre and Edward to a greater or lesser extent experience. Deidre has set up her life according to a commitment to mathematics; through her actions she furthers the progress of mathematics. That is to say, in Frankfurt's terms, Deidre appears to love mathematics as she consistently acts out of a concern for mathematics. Yet her heart is no longer in it. Her existential unease cannot be captured, let alone elucidated, without taking this loss of positive affect regarding mathematics into account.

Inherent to the term 'relation' is the idea that it is something between two entities. We relate ourselves to others who are fundamentally important to us. Exactly because they are so important to us, it matters to us who they are, existentially speaking and otherwise. Crucially, it matters to us what they think of us, whether we matter to them and what their affective-volitional relation to us is. That is to say, our significant others have an influence on the relation in which we stand to them. In contrast, Frankfurt does not take the beloved and the beloved's relation to the lover into account when describing what loving is or should be. When beloveds are young, helpless infants, Frankfurt's favorite example (2004, 43), this may make some sense. Yet insofar as friends, family members, partners and other beloveds are autonomous beings, it seems strange not to take their behavior into account when determining our own stance towards them (cf. Ebels-Duggan 2008). That is to say, Frankfurt's view does not seem to capture the phenomena accurately: how our significant others relate to us has an impact on how we relate to them. A relation view is able to accommodate this, and the affective-volitional ambiguities it may entail.

The relation in which our significant others stand to us may be very different, in its affective-volitional character, than the relation in which we stand to them. There is not one relation between us and them, but two: one relation from them to us,

constituting them; and another relation from us to them, constituting us. They may be incredibly important to us and we may love them wholeheartedly, whereas we are not quite as important to them, and their relation to us is more ambiguous, involving boredom, for example. Or there may be one relation present only, as when fans adore a pop star or writer and paper their bedrooms with posters or buy all their books, watch all the videos that can possibly be found of them online, day dream about meetings with them and feel happy while daydreaming —without the adored pop star or writer even knowing of their existence.

In sum then, existential selfhood is structurally characterized by affective-volitional relations to people and projects of fundamental, personal importance. These relations capture the distance that remains between us and even our most beloved others. Although we may not be able to decide at will to start or stop loving someone, we have some influence over the relation in which we stand to them. Compared to Frankfurt's account, this relational view is better able to accommodate our significant others' influence on our emotional lives, the ambiguities present in everyday loving relationships and the intersubjectivity of love. What is more, the relation account brings into view a ubiquitous practical problem: how should we relate to the people that are existentially important to us? We question the self-constituting relations into which we stand to our significant others. It is through replacing Frankfurt's 'identification' with my 'relation' that integration with some of the insights we may learn from Charles Taylor and Søren Kierkegaard becomes possible.

3. 'Relations': reflexivity and the loops that we are

Harry Frankfurt emphasizes how we are shaped by what we love, which is to a large extent not up to us. In contrast, Søren Kierkegaard, particularly his pseudonym Anti-Climacus, emphasizes how we constitute ourselves by taking a step back from our automatic ways and consciously relate to who we are. Narrativists such as Charles Taylor emphasize how our narrative self-understanding shapes us. Meanwhile, empirical disciplines such as cognitive neuroscience and psychology remind us how much of our behavior happens on auto-pilot, so to speak. The situations in which people stop doing so for a moment and reflect on themselves are often situations when there is something at stake: when their actions or emotional reactions surprise them, when their existential identity is no longer self-evident, when important decisions have to be made, or when persistent feelings of dissatisfaction lead them to question what is truly important to them. Through reflection, people form an explicit understanding of themselves in terms

of what matters most to them and how they relate to what matters most. What does such self-reflection involve?

First of all, people interpret their actions and experiences and formulate a piece of self-understanding that has cognitive content. They form an understanding of the relations that shape them. They conclude that someone is more important to them than they would have thought. Or they realize that their relationship to them is not as straightforward as they may have thought. They thus establish something about the particular existential selves that they are. The persistent dissatisfaction that they experience becomes a little less inchoate as they form an understanding of it.

People scrutinize their behavior and experiences to see what these may have to tell them about who they are and in what direction they are becoming. In turn, the understanding they thus form influences their behavior and experiences, which is to say, their selves. 'Why do I act like I do?' we can imagine Bob asking himself. 'Why do I react like I do to my dog in the ways that I do; why, for example, do I experience such flaming anger at my dog so often?' It matters immensely whether Bob formulates an answer along the lines of 'because my dog is an unruly mongrel' or 'because I have a tendency to pick fights with whomever is close to me. I did the same with my wife when she was still alive.' Self-understanding's cognitive content thus has important consequences for the affective-volitional relations it reflects on. If he decides that the dog is the problem, Bob is bound to dislike the dog even more and may decide to get rid of his dog. Alternatively, if he thinks the issue lies with himself, he is bound to empathize more with the dog from now on and seek to improve his own behavior. To be clear, I do not argue that one or the other option is better, but only wish to point out structural characteristics of these phenomena: when people form an understanding of the existential aspect of their selves, this changes their selves, as it changes the relations in which they stand to their significant others.

As another example, say we stand in an unproblematic, loving relation to our father. Yet somehow we become convinced that our relation to him is much more ambiguous. Maybe we read that grown-up children always need to break away from the dominating presence of their fathers and that they experience resentment for never being fully able to do so. Or maybe some friends of ours project their own troubles with their fathers onto our situation. In any case, when we meet our father, he asks us to help him with some little chore. Right before we almost automatically do so, we think to ourselves that we need to break away from our father's dominating presence. Therefore we tell him that we do not want to help. Our father looks surprised and a little hurt. This makes us feel bad about refusing to help, which makes us experience resentment at our father for making us feel bad for refusing. Et voilà, our understanding of the relation in

which we stand to our father has now, like a self-fulfilling prophecy, lead to ambiguities in the relation in which we stand to our father that forms part of our existential self.

Self-understanding and selfhood form a constant loop. Both Taylor and Kierkegaard's pseudonym Anti-Climacus explicate how self-understanding changes selfhood. Taylor writes that insofar as people can embrace what they find themselves to be, the motivations that result from the relations in which they stand to significant others and significant projects are likely enhanced (1989, 96). People may act more easily on these motivations and gain a stronger sense of purpose through that. They may gain self-confidence as they now understand themselves better and like what they see. Vice versa, inner conflict may result when people find that they value different things that clash, or value something that they wish they would not value (1989, 107).

When what people find themselves to be is not in line with what they would like to be, their dissatisfaction with themselves gets intensified, at least at first. Anna does not admit it into so many words to herself, but she cannot stand herself for failing to experience positive connections to her husband and her work. If she starts to reflect on this, however, it is likely to intensify her dissatisfaction with herself at first. After all, Anna then has to face the disturbing fact that she currently cannot experience any love for her husband, nor for her line of work. Of course this may provide the impetus for her to try to change her situation. It need not, however, as Anti-Climacus explicates. Anna may use all her powers to deny her discovery, to others and also to herself. Anti-Climacus writes of the "interplay between knowing and willing" or what I would call the cognitive and affective-volitional dimensions to existential self-understanding. Using Anti-Climacus' words, we can see how for a while Anna tries "to keep [her]self in the dark about [her] state through diversions [...], through work and busyness as diversionary means, yet in such a way that [s]he does not entirely realize why [s]he is doing it, that is to keep [her]self in the dark" (1980 [1849], 48).

I would like to add that when people cannot embrace what they think they are, they may not just try to forget about it, but may also actively deny it through forming a different story about themselves. That is to say, besides effects of the cognitive content on the affective-volitional relations (which Taylor focuses on) and besides the reverse effect in the shape of people trying to repress cognitive content because they do not like it (which Anti-Climacus speaks of), people may also *alter* the cognitive content of their self-reflections, to be more in line with what they would like to be. That is to say, people may not just suffer from self-induced oblivion with respect to a part of themselves, they may also flat out delude themselves with respect to who they are. After all, self-reflection is an activity that involves the interpretation of your own behavior and experiences. And people come up with interpretations serving all sorts of purposes.

Edward, for example, strives to be a dignified butler. In his opinion, one of the things a dignified butler should definitely *not* do is fall in love with the housekeeper. To the readers of Edward's story, his emotional reactions to Ms. Kenton and his behavior towards her clearly signal love. Yet he interprets them time and again as indicative only of a professional relation between a butler and a housekeeper. In the short run, he manages to quiet any lingering unease he may have about not living up to his own ideals. In the long run, he lives a life that is disconnected from important personal feelings. Whether this is bad or not is not the issue here; I want to point out structural characteristics at play in existential selfhood. One of them is that the cognitive understanding we form of our relations to significant others interacts with those relations in both directions. Plus, when we do not like what we find ourselves to be, we may not just repress cognitive content, but also replace it with content that is more in line with how we wish things were.

All in all, we are implicated in our selfhood. We are not just passive victims or passive beneficiaries of the circumstances that befall us, but influence the course of our lives. Implicit in any understanding of the particularities of our selfhood is the discovery *that* we are existential selves. We discover that people and pursuits can be of utmost importance to us, and that our relations to them have an impact on how we experience things and what we are inclined to do. We also implicitly discover that we are capable of self-reflection, of distancing ourselves from what we are automatically inclined to do and experience. It is *in our power* to ponder what else could be important to us and how else we could relate to what is important to us. We discover questions to ask ourselves in this respect. At a distance from our automatic, unreflective tendencies, we can imagine that we could lead our lives differently. The question may arise whether the way we currently are relating to our lives and what is vital to us in them, is the best way, or at least good enough, at least for us, at least for now. We can make plans on how to change ourselves. We can decide to alter some of the circumstances of our lives to make it likely that we change as we envision. In sum, we discover that it is to some extent up to us to determine who we are.

When people live through existential unease, as Anna, Bob, Clemens, Deidre and Edward all do to some extent, they may decide that it is time to take action. Again, I do not argue that they should; this is not a normative argument regarding how people should deal with the existential aspect to their lives. Instead, I explicate the structure of existential unease and the corresponding ways in which it may be relieved, at least to some extent. People may try to alter something in the relations to significant people and significant pursuits and/or in the understanding that they hold of what these relations

should be like. They may attempt to bring who they are and who they would like to be more in line with each other.

To start with the latter, people can ask themselves what it is they think they would like to be and whether they actually agree that this is what they would like to be. They may decide that what they think they would like to be makes unrealistic demands on them, for example. Deidre may feel that she should be a mathematician as her entire family works 'with numbers'. Disavowing this ideal may release much existential tension and give her the space to act on what she is actually attracted to. Anna may have a sense that she should relate to husband and children in a perfectly loving way all of the time. Realizing this a rather extreme demand and replacing it with a more moderate ideal may take away some of the unease she experiences for being unable to do so at present. The particular sense people have of who they should be is unlikely to go away immediately and may never leave them entirely, especially if it has been engrained in them from early on in their lives. Yet, it does not hold the same force over them when they consciously reflect on it and reject it as a valid ideal.

Alternatively, people may tackle the relations in which they stand to their significant others, in the hope of changing the actual behavior and experiences they have. Whether at New Year's or otherwise, people vouch to change the way in which they relate to what is most important to them. People may make resolutions to become more sociable and start immediately by calling their friends and family for a chat and to make appointments for visits. People may try to talk things over with someone who is dear to them when their relationship has soured, in the hope of improving it again. People may try to contribute to gaining or losing a few significant others, by taking a dog or signing up or a dating site on the one hand and on the other hand by leaving a job or leaving a lover. Explicit reflection allows people to check where they stand with respect to becoming who they would like to be.

While Frankfurt emphasizes that we are not free to establish the outlines of our existential selfhood, Anti-Climacus emphasizes our freedom to constitute ourselves. That is to say, he acknowledges that we have not established ourselves and therefore have to deal with possibilities and limitations that are not of our own making. In light of what I just wrote, I could give the example of 'that which we think we would like to be': it often consists, at least at first, of an internalization of what our social environment holds up as exemplary, much in line with the 'goods' and 'supergoods' of which Taylor speaks that stem from our social environment. We do not choose the environment into which we are born, yet cannot escape its formative influence. This is but one of the ways in which we do not establish ourselves and that we have to deal with. Yet Anti-Climacus thinks that exactly in our dealing with ourselves, in our relating to ourselves, we are entirely free. He holds us entirely responsible to deal with ourselves in ways that are, in

his eyes, poorer or better. In light of empirical research, however, it seems that we are creatures of habit also in our ways of relating to ourselves and to others. Attachment theory posits as much and psychological research appears to back this up (Cassidy and Shaver 1999; Holmes 1993). Normative theorists may decide to take this empirical fact into account. Importantly, this requires an additional decision, to adjust one's views on how people should relate to their significant others depending on how they, on average, are able to relate to them. This decision is one that requires argument and not every normative theorist may be willing to make it. I think Anti-Climacus is unlikely to change his normative views in light of what people generally are like, deeming any possibility, however low in likelihood, enough possibility to ascribe responsibility. In terms of the phenomena, when we consider that humans are able to distance themselves from themselves, developing an understanding of themselves that influences the direction in which they develop, we may want to add: to some extent at least.

4. Questions at stake in existential self-understanding

In this study, I have chosen to approach 'existential selfhood' and 'existential self-understanding' in a bottom-up fashion. Instead of deciding at the outset how to conceive of these notions, I started with a few examples of phenomena in which they are at play. I then examined the contributions CNS literatures concerning love and self-reflection and a few representative philosophers have to make to our understanding of the existential aspect of the human condition. CNS contributes to our insight into the neural enabling conditions underlying the possibilities for finding particular people and projects fundamentally important, for relating to these significant others in various ways, and for reflecting on these self-constituting relations. Philosophers elucidate existential selfhood at the level of lived experience and articulate structural characteristics of existential phenomena. Often, they do so by relating their own analyses of those structural characteristics to those of others, unearthing assumptions embedded in other philosopher's views, building on these where possible and criticizing them where necessary. In this chapter, I aimed to do as much, synthesizing important insights by Harry Frankfurt, Charles Taylor and Søren Kierkegaard, building on the analyses and criticisms I provided of their views in chapter 3. This is not just relevant in relation to existing philosophy of existential selfhood, but first and foremost as it elucidates the phenomena: the behavior, experiences and problems that we encounter as the existential aspect of our lives and that are exemplified in the stories of Anna, Bob, Clemens, Deidre and Edward.

In brief, I argue that existential selfhood is formed by the relations in which people stand to people and projects that are of fundamental significance to them. These relations are affective-volitional in character, which is to say that there is an inextricable connection between how people feel about their significant others and how they are motivated to act towards them. As humans are able to reflect on who they are, they are able to form a cognitive understanding of the affective-volitional relations that constitute the existential aspect of themselves. This understanding may influence those relations. Yet when people do not like what they understand themselves to be, they may also repress their understanding or adapt it so that it is more in line with who they would like to be. When people have existential problems that they wish to solve, they may try to bring who they are more in line with who they would like to be. That is, they may address their relations to significant others, aiming to change the actual behavior and experiences they have. They may also scrutinize their ideas about who they would like to be, to see whether they in fact endorse them upon reflection. Whatever the result of these reflections is, it is likely to change them and these changes have therefore been influenced by themselves. Yet people also change in ways that they have little influence over, for example when their significant others change or the relationship to significant others changes due to changing circumstances. At some point, all the changes lead people to wonder anew who they are in terms of what is most important to them and how they relate to what is most important to them. And so on: the ongoing loop of selfhood.

This view on existential selfhood is indebted to views by Harry Frankfurt, Charles Taylor and Søren Kierkegaard's pseudonym Anti-Climacus. Particularly, I draw on Frankfurt's ideas about love and his explications of the structural connections between love and selfhood, on Taylor's views of what articulating an understanding of oneself does to selfhood and on Anti-Climacus' writings on the relations that constitute selfhood. Yet my view is different from Anti-Climacus' view, as Anti-Climacus espouses a radical conception of selfhood in which selfhood is constituted by people's relations to themselves and to God. He acknowledges that people ordinarily often think of themselves in terms of their relations to people and pursuits that are important to them, but does not think this is true selfhood and therefore does not explicate how these relations may constitute selfhood, as I do. My view is different too from Taylor's views, which revolve around values that are socio-culturally shared and therefore do not individuate people in the way that existential selfhood does. Last but not least, my view is different from Frankfurt's, who describes selfhood in terms of volition, as consisting of an acceptance of the significant other's interests amongst our own, as our own. I argue that my ideas about the *affective*-volitional relations constituting selfhood are better able to clarify the influence our significant others have on our emotional lives and to

elucidate the ambiguities that are often present in human relationships. What is more, I think a relational view does more justice to the intersubjectivity of loving and thus brings an important practical problem in view: how should we relate to our significant others? All in all then, my view on existential selfhood is a view on how various strands within existing philosophical thought may be integrated and that future philosophy of existential selfhood may build on.

As philosophical analyses often do, my analysis clarifies the questions that are at stake in existential self-understanding. I can outline them now in an order corresponding roughly to the different sections of this chapter. First, who and what is of fundamental importance to us? That is, what affects us profoundly? Where do our strongest motivations stem from? Second, how do we relate to whom and what is so important to us? That is, what are our feelings for our significant others? How do we feel about the projects we have invested some of our identity in? What are we naturally motivated to do with respect to them? And how do our significant others relate to us? Third, how does the understanding that we thus gain compare to ideas we already held about ourselves? How does it compare to our ideas about what we would like to be? In fact, what are our ideas about who we would like to be? And do we actually endorse them now that we reflect on them? Is there something we can do to our lives and ourselves to live more in line with who we would like to be? Is there something we can do in the hope of changing our significant others' way of relating to us? Finally, there is a question to be added on the basis of the previous chapter, particularly relevant in cases where we appear fundamentally, pathologically unable to relate to others or to self-reflect as we would like to. Has some enabling condition gone unmet preventing us to love or self-reflect? Is there something we can do to enable us to love and self-reflect adequately again?

The clarification of these questions may be relevant in several ways. Where it comes to CNS, these questions may be front-loaded into CNS experimental design, if neuroscientists would want to extend the range of their research to include self-reflection on the existential aspect of the self. Where it comes to philosophy, the work I do here may serve as preparatory work for the development of a normative view on existential selfhood. Philosophers who wish to develop such a view may focus on any of the above questions and have to take into account the various other structural dimensions to existential selfhood that I outline too, I would argue. Particularly, my analyses brings into light a practical question that has not received much attention so far. Susan Wolf and Harry Frankfurt have argued about who and what we should love and on what basis we should decide this (Wolf 2002; Frankfurt 2002). My analysis points to the relevance of a different question: how should we relate to the people and projects that are fundamentally important to us, that is, how should we deal with them?

Finally, the clarification of these questions may be relevant to all of us, including Anna, Bob, Clemens, Deidre and Edward, where we are interested in understanding our own existential situation better. After all, the understanding we hold of ourselves is often hazy, and existential dissatisfaction often inchoate. As I said in chapter 3, philosophy does not provide therapy for people's troubles, existential or otherwise, as it does not take people's personal circumstances into account. At the end of this chapter, we therefore do not know how Anna, Bob, Clemens, Deidre and Edward should resolve their situations. However, the questions that Anna and the others may ask themselves, or that their significant others or therapists may ask them, are hopefully clearer now than before.

Philosophy may contribute to our understanding of existential selfhood and existential self-understanding by analyzing the structural characteristics of phenomena as well as structural connections between them. In the process, they often point out phenomena that we glance over in daily life. Philosophers also unearth assumptions present in ways of thinking about existential issues and critically question these assumptions. Furthermore, philosophers may develop normative views on how to lead a good human existence. In effect, philosophy clarifies the questions that are structurally at stake in the existential aspect of the human condition. It thereby clarifies the questions that all of us may ask ourselves and each other when we wish to understand ourselves better.

Conclusion

How should we understand the existential aspect of the human condition? And what can we learn from cognitive neuroscience (CNS) and philosophy regarding it? Cognitive neuroscience has not entered the academy all that long ago. It is gradually moving into terrain that was once thought to belong exclusively to the humanities. Cognitive neuroscientists study love, self-reflection, moral judgment, and religious experience, for example. Many think CNS is bound to increase our understanding of the human condition (e.g. Bechtel, Mandik, and Mundale 2001; Churchland 2008; Gazzaniga, Ivry, and Mangun 2002). Regularly, these claims are framed in opposition to philosophy. Cognitive neuroscientists perform experiments; philosophers rely on their intuitions and interpretations. CNS is therefore able to contribute to human self-understanding on a more reliable basis than philosophy can, the argument goes (Churchland 2008; Gazzaniga, Ivry, and Mangun 2002). On the other hand, one can find thinkers who are skeptical that CNS has anything truly interesting to offer to human self-understanding (Bennett and Hacker 2003; Tallis 2011). With this study, I aim to clarify three issues at stake in these debates: first, the questions that CNS experiments address and the status of its answers given CNS' research process; second, the understanding philosophy provides and its status, given its research process; third, the understanding that we search for when searching for existential self-understanding.

1. CNS and philosophy: coexistence and collaboration

Cognitive neuroscience is an experimental discipline, yet as is the case with all experimental disciplines, interpretation plays an important role at several moments in its research process. First, tasks are developed and participants are recruited with pragmatic considerations in mind and with an understanding of the phenomenon under study in mind. Second, researchers need to interpret neural data in terms of what it has to say about the phenomenon under study. Researchers' existing understanding may influence their interpretation here too, as for most neural activity, several (if not many) types of tasks are known to elicit it and there are thus many possible interpretations. For example, certain researchers hold a theoretical conviction that love is a motivation

rather than an emotion and then interpret particular neural activity as representative of motivational processes, not emotional ones (Aron et al. 2005). Other researchers find neural activity that is difficult to interpret and decide to draw on socio-cultural ideas: this neural activity may represent mechanisms by which love makes blind, they say (Bartels and Zeki 2004). To an uncritical eye it may seem as if neural evidence is found for love as motivation or for love making blind, whereas actually these researchers have interpreted their data in terms of their existing understanding of love.

CNS as a field has put its hope for progress in the aggregation of large numbers of studies, for example in review articles and databases. Several hurdles are in the way here, however, that cognitive neuroscientists mostly seem to be unaware of. First, many studies that are labelled the same are in fact not directly comparable, as they employ different tasks. Vice versa, many studies that are labelled differently and are thus not categorized together in fact employ the same task and thus should be compared directly. Second, just as data aggregations rely on labels that researchers have come up with, databases in particular aggregate the interpretations researchers give of their data. They take these interpretations and the pre-existing built-in understandings for granted instead of questioning them. Third, this is particularly detrimental given that research practice contains incentives to repeat earlier interpretations. Researchers employ tasks that previous studies have also employed. They interpret neural activity along the lines that earlier researchers have also suggested. A body of research may compile that seems to support a particular understanding of love or self-reflection, whereas what is in fact compiling is a body of data that researchers interpret along the same lines over and over again, even though other interpretations could be just as valid.

The lack of self-critical attention to interpretative moments in CNS' research process is an important barrier to progress for the field. Conceptual reviewing work could be an important tool to implement self-critical reflection in CNS, and I will come back to it shortly. Let us say that it will be implemented and that CNS develops greatly on all fronts over the years to come. What can CNS contribute to our understanding of the existential aspect of the human condition then? Foremost, CNS provides insight into the brain-in-action. Through it, we gain an understanding of the neural enabling conditions of ordinary behavior and experience. This is relevant for self-understanding: in dealing with ourselves, we need to deal with the fact that we are embrained beings, so to speak. If we are interested in being able to love and to self-reflect, we should make sure enabling conditions are met. Also, we have to reckon with automatic tendencies that drive much of our behavior. Nevertheless, insight into the neural enabling conditions does not provide insight into the way in which we relate to our significant others, or into the contents or quality of our self-reflections. Once enabling conditions are met, further insight into them does not increase understanding of the phenomena

they enable. Furthermore, inferences about behavior drawn on the basis of neural data may increase our confidence in particular psychological theories over others, when they triangulate with other types of data. However, I argue that direct evidence is to be preferred over indirect evidence and that future reverse inferences based on neural data therefore cannot trump data from behavioral experiments when it comes to understanding the psychological processes underlying behavior.

How about philosophy then? What does philosophy contribute to our understanding of existential self-understanding? Philosophers working on topics that pertain to existential self-understanding do several things. They aim to describe in articulate words phenomena that are part of our lived experience yet that we may glance over in daily life. Philosophers try to tease out structural characteristics at play in these phenomena, as well as structural connections between them. They clarify the questions at stake in existential self-understanding. Philosophers also analyze common ways of thinking about existential selfhood and the assumptions underlying ways of thinking. They may criticize particular assumptions and propose alternative ways of thinking about phenomena, arguing why it is that one way of thinking is to be preferred over another. Their clarifications are not always easy to understand; upon reflection lived experience often turns out to be more complex than initially thought. Yet philosophers try to clarify people's lived experience, rather than underlying enabling conditions of lived experience. Furthermore, they often develop views on how people should deal with existential situations, that is, they put forth normative views on existential selfhood.

In doing so, philosophers rely heavily on interpretation. They aim to be self-critical throughout. Once they have formulated an idea about a phenomenon, they try to clarify what they mean by clarifying the terms they use to describe it. They try to clarify and question the presuppositions on which the idea rests and the implications it has. They question these too and try to think of examples that would disprove their take on the phenomenon. So do their colleagues. If a counterexample indeed can be found, then the search for the basis on which the examples differ serves to hone the idea more precisely. This system is not perfect, as phenomena may be overlooked. Especially phenomena that easily go under the radar of conscious reflection may be overlooked, such as automatic behavior, or also phenomena that are not on the minds of the somewhat privileged people engaging in academic philosophy. When philosophers are not self-critical and do not listen to the criticism that their colleagues come up with, the quality of their work suffers. Although the system is not watertight, self-critical reflection is still the best way to deal with interpretation. In fact, one could argue that insofar as people attempt to delve up assumptions underlying interpretations and to

critically reflect on them, they are engaging in philosophy. Systemically, academic philosophy promotes self-critical reflection.

All in all then, cognitive neuroscience and philosophy mostly engage in entirely different things. To speak of an opposition between them therefore does not make much sense. Cognitive neuroscientists spend the overwhelming majority of their time and energy on studying the brain, particularly the brain in action. The brain is still largely a terra incognita; studying its involvement in our behavior and experiences is a fascinating quest in itself, even if it does not often result in insight that people may use to understand themselves better. Philosophers need not be defensive towards or dismissive of cognitive neuroscience therefore, for it should not be understood as some sort of pseudo-philosophy. Vice versa, philosophy does not equal pseudo-CNS, far from it. It is not some imprecise, speculative forerunner of experimental work. Mostly, philosophy of existential selfhood addresses questions that cannot be approached empirically and do not pertain to the enabling conditions underlying lived experience, but to lived experience itself.

Where there is some overlap between the disciplines, they can and should work together rather than in opposition to each other. Parts of philosophy that do in fact develop or rely on ideas about the enabling conditions to human behavior and experience may be informed by CNS. CNS may constrain theory formation about mind-brain relations, for example. Although it cannot tell us how we should conceive of the relation between mind and brain, its practice may point out untenable positions, such as those that deny consistent correlations between neural activity and behavior. Also, CNS, in conjunction with psychology, reminds us of the automatic tendencies that ground much of our behavior and the (biological) factors that influence our actions also in circumstances where we may not have thought that they do.

Cognitive neuroscience on the other hand may benefit from interaction with philosophy where conceptual issues are at stake. In this study I have discussed ways of improving the interpretation of results but I focused first and foremost on the conceptual implications of task development. I propose conceptual review as a way to move the field forward. Conceptual review involves setting up an inventory of the choices researchers make during task development and teasing out the conceptual implications of these choices. This allows one to say what is being investigated precisely and thus to analyze the validity of the task, the limitations that it has, whether there are any theoretical or socio-cultural biases embedded, and how the task relates to other tasks that come with their own limitations. In turn, this helps to provide suggestions on how to move the field forward. Certain practices may have to be dropped. Focused diversification of tasks is made possible. Last but not least, conceptual review facilitates

ordinary review of CNS literatures and the setting up of databases, as it aligns tasks and labels that are currently inconsistently connected. In current CNS of self-reflection, for example, two types of task are employed, yet at least nine different labels are used to refer to them, with certain labels used to refer to either task. Conceptual review is one way in which philosophy could be *for* cognitive neuroscience, and I think it would facilitate other ways. It could bring to light concepts that require analysis, for example, or point out areas where it would be useful to front-load phenomenology into research design.

In the introduction, I quoted a textbook used to teach the basics of cognitive neuroscience to students throughout the world. Its first and second editions introduce CNS as a scientific approach to "big things like the meaning of life, or the meaning of meaning" (Gazzaniga, Ivry, and Mangun 2002, 1). It contains a quote by "the brilliant philosopher Søren Kierkegaard" in which Kierkegaard dismisses the idea that a natural science could study consciousness. The authors of the textbook beg to differ and go on to position cognitive neuroscience in opposition 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 2002, 2). At the end of chapter 4 we saw that the third edition of the textbook does not contain the wild promises regarding CNS' contributions anymore, though the opposition to philosophy remains unchanged (Gazzaniga, Ivry, and Mangun 2008). In the most recent, i.e. fourth, edition of the textbook, the explicit positioning of cognitive neuroscience in opposition to philosophy is no longer there. What remains is this: "To understand how biological systems work, a laboratory is needed and experiments have to be performed [...] This approach is known as the scientific method, and it is the only way that a topic can move along on sure footing" (Gazzaniga, Ivry, and Mangun 2014, 5).

I think the changes from edition to edition are representative of a broader trend amongst cognitive neuroscientists. There appears to be a growing awareness that the field cannot live up to the high expectations with which it was introduced in the beginning of the so-called decade of the brain (cf. Tretter and Kotchoubey 2014; Jarrett 2015). This does not make my analysis superfluous, however, quite the contrary. Mostly, it is not clear to cognitive neuroscientists, nor to lay people drawing on cognitive neuroscience, why CNS cannot fulfill the high hopes. Many may be inclined to think that it will simply take longer than expected for CNS to overthrow our existing ideas about ourselves, without doubting CNS' potential to do so. Some neuroscientists realize that there are fundamental obstacles in the way and may even be interested in philosophy because of it. Yet it does not always become clear to them what philosophy

has to offer. It does not always become clear to them either on what basis to trust philosophical insights, given that these have not been obtained by means of experiments.

What I hope to have shown in this study is that although it is indeed true that one needs a laboratory and experiments to study how a biological system works, experimental work necessarily includes moments of interpretative choice. CNS' footing, as the textbook has it, may thus be a little less sure than some researchers and lay people like to think, and the meaning of CNS results less univocal. The best way to proceed is to acknowledge moments of interpretation and self-critically scrutinize the choices that are being made. Philosophy may be of help here. What is more, I hope to have shown that not any topic can be approached through experimental work. Conceptual questions cannot be decided by means of experiments, for example, nor can normative ones. The 'perspective' that philosophy offers and the 'eloquent and fascinating ideas' it puts forth mostly cannot be put to the test via the scientific method; they are of a different kind than the hypotheses whose truth cognitive neuroscientists strive to establish.

What I also hope to have done then, is to have underlined the positive challenges that the presence of CNS in the academic landscape presents for philosophers. There is work to be done. First, there are contributions to be made to CNS research, for those who are willing and capable to take that research seriously and do the work. This study should help them on their way. Moreover, there is work to be done in terms of communicating philosophical research to audiences beyond fellow philosophers. To some extent any professional field is a bit of a bubble and academic philosophy is no exception. Within the bubble, it may seem enough to hand-waivingly dismiss CNS as irrelevant to the truly interesting questions about human existence, as I sometimes encounter philosophers doing. Outside of the bubble, however, neuroscience is being appropriated in fields as diverse as law, public policy, education, parenting, economics and marketing (O'Connor and Joffe 2013), often in unwarranted ways (Satel and Lilienfeld 2013). Neuropopularization reaches large audiences. It is vital to communicate with cognitive neuroscientists and others about what philosophy is and can do. The confrontation with CNS research brings out the urgency of philosophical anthropological work, for example philosophical work investigating what is at stake in self-understanding. What is more, it brings out the urgency of philosophy as an integrative endeavor: philosophers study the types of questions that various academic fields pursue and how these relate to each other. With this study I hope to have contributed on both fronts. That is, I hope to have shown the relevance of philosophical anthropological work, as well as the relevance of philosophical reflection on CNS' and philosophical anthropology's potential contributions to our understanding of the human condition.

2. Existential self-understanding

So where does that leave us in terms of understanding existential self-understanding? What insight have we gained regarding the existential aspect of the human condition, as exemplified by Anna, Bob, Clemens, Deidre and Edward? Throughout this study the phenomena at stake in existential selfhood and existential self-understanding have been pointed out to us. Structural connections between what may have seemed to be disparate phenomena have been clarified. In the process, I have also teased out and criticized assumptions that fellow philosophers make regarding existential self-understanding.

The people and projects that are of fundamental importance to us have an important bearing on our emotional and practical lives: what they do and what happens to them does not leave us cold and we are naturally motivated to act so as to influence their fate. Harry Frankfurt is the philosopher whose thought pertains most directly to these issues. Contra Harry Frankfurt, I do not think our significant others and significant projects shape us through a direct positive identification with them and their interests. Rather, the affective-volitional relation in which we stand to them determines how they affect us emotionally and practically. When we absolutely hate them, they do not leave us cold, but only in the sense that we derive some satisfaction from their failings and dissatisfaction from their successes. We may be motivated to act so as to hinder them. On the other hand, when we love those who are important to us in perfectly unselfish ways, we rejoice when they thrive and are fully motivated to act so as to increase chances that they will. Most human relations fall somewhere in between these two extremes. Real-life relationships between romantic partners, parents and children, siblings or friends are hardly ever entirely devoid of some ambiguity, nor are relations from people to the projects they have invested their identity in. This also accounts for the existence of a ubiquitous practical problem: how should we deal with our significant others? Frankfurt's account does not focus on this question and cannot accommodate the fact that we take our significant others' stance towards us into account when answering it. In contrast, a relational account does justice to the prominence this question has in our lives.

Two seemingly opposed strands of thought run through philosophies of existential selfhood. Some philosophers emphasize that the shaping of our existential selfhood happens to us. Others emphasize the influence we have over shaping ourselves. I think a relational account of existential selfhood provides a good basis to reconcile these two ideas and hope that future work will flesh it out further. We discover who and what is of fundamental importance to us more than that we decide it. Yet we influence the relations in which we stand to them and that shape our emotional and practical lives.

Even if we are to a large extent not free to determine what is of fundamental importance to us, having fundamentally important people and pursuits in our lives is liberating in many ways. It gives us a sense of direction as it provides us with a fundamental set of motivations and thus a basis for choice. What is more, if we are able to contribute positively to the lives of those we love, this may give us a sense of purpose. We may experience our lives as meaningful instead of meaningless.

Existential self-understanding is inherently personal. It is not about people in general, but about Anna and the particular individual situation she finds herself in. Unlike biographers and therapists, philosophers and cognitive neuroscientists strive for an understanding of human beings, rather than for an understanding of one particular human being and then another and then another. Therefore, this study does not provide answers regarding what Anna, Bob, Clemens, Deidre and Edward should do about their situation. Such answers require insight into the details of their particular situations, which neither philosophy nor CNS are focused on. What we have gained with the help of philosophy, however, is a clearer insight into the questions that people may ask themselves to clarify their existential situation to themselves. They may ask themselves: who and what is of fundamental importance to me? What are the relations in which I stand to my significant others and significant projects like? And how does this compare to how I would like to relate to the people and projects that are important to me? If there is a distressing mismatch between who I think I am and who I think I would like to be, can I change either side of this mismatch? That is to say, is there something I can do to try to change the relation to those who are of fundamental importance to me? Do my ideas about who I would like to be stand up to scrutiny now that I reflect on them? If somebody is unable to care about anything or unable to self-reflect altogether, an enabling condition for existential selfhood and existential self-understanding has gone unmet. If this is a neural enabling condition, CNS may provide insight into how to improve the situation. Yet information about the neural activity correlating with thinking about significant others is not going to provide insight into the relations people have to their significant others, nor is it going to give them a sense of direction regarding how to try to influence their existential situation. For that, they need to ask themselves the questions I just mentioned.

The understanding people form of themselves in response to questions such as these influence the direction in which they are developing. Existential selfhood and existential self-understanding form an ongoing loop. More generally, human beings are becoming beings, slowly changing throughout their lives. Long periods of time may pass in which people are relatively at peace with the relations in which they stand to their significant others, and with the fundamental motivations and emotional responses that are the result of these relations. Mismatches between who one is who one would like to

be are bound to occur, but they may remain implicit for a long time. At some point, however, a kind of nagging dissatisfaction may lead one to question anew: Who am I? What is truly important to me? And how do I relate to what is truly important to me? The restlessness that is explicated by these questions is part of the human condition. It is a tolerable restlessness most of the time. It is what makes human beings striving beings. The quest for self-understanding, existential and otherwise, will be ongoing as long as there are human beings alive.

Appendix A

Study	Selection criteria participants	Participants	Stimuli	Instructions	Control task	Check after experiment
Acevedo et al. (2012) Long-term romantic love	Flyer, newspaper ads, word of mouth: 'Are you still madly in love with your long-term partner?' Phone screening: Relationship length >10 years Monogamous heterosexual relation Feelings of intense romantic love	10 ♀ / 7 ♂ equals ± 60% of original respondents 39-67 years old (53 years) 'being married': 10-29 years (21 years)	color picture partner	'think about (non-sexual) experiences with person'	Same instructions. Pics of: close friend highly familiar acquaintance low familiar acquaintance (all same sex and ± same age)	Participants described thoughts and feelings while viewing stimuli
Aron et al. (2005) Romantic love	Flyer, word of mouth: 'currently intensely in love'	10 ♀ / 7 ♂ 18-26 years old (20.6 years) 'being in love': 1-17 months (7.4 months)	picture beloved 30 sec	'think about events with person that were pleasurable, but not sexual'	Same instructions. Pics of: familiar acquaintance (of same sex and age) countback task	Interviews to check whether participants followed instructions and what they thought about during scan.
Bartels & Zeki (2000) Romantic love	Posters, via the internet: 'truly, deeply and madly in love' Written statement describing how much they are in love + interview	11 ♀ / 6 ♂ out of 70 original respondents, 75% of which female (write B&Z) 21-37 years old (24.5 years) Relationship: (2.4 years)	color picture partner 17.36 sec	'view the pictures, think of the viewed person and relax'	Same instructions. Pics of: three friends (of same sex, similar age, and known for at least as long as partner)	Participants rated felt love and sexual arousal (1-9) when viewing partner or friends.
Bartels & Zeki (2004) Maternal love	Posters in nurseries: being mother	20 ♀ 27-49 years old (34 years)	picture own child 15 sec	'view the pictures and relax'	Same instructions. Pics of: other child (known to participant, same age as own child)	Participants rated intensity of eight different feelings for

Study	Selection criteria participants	Participants	Stimuli	Instructions	Control task	Check after experiment
		'age child': 9 months – 6 years (24.4 months)			best friend acquaintance	each person viewed.
Beauregard et al. (2009) Unconditional love	Directors of l'Arche residential communities where people with intellectual disabilities live together with assistants helped recruit those assistants: 'with a very high capacity for unconditional love'; who understand the meaning of 'unc. love'; and find work at l'Arche very gratifying.	9 ♀ / 8 ♂ 20-63 years old (36 years)	pictures of unfamiliar children and adults, all with intellectual disabilities 9 sec	'self-generate a feeling of unconditional love towards the individuals depicted'	Similar pictures. Instructions:: 'simply look at the individuals depicted'	After each block of 4 pics, participants rated how much unconditional love they felt (1-4). Post-scan interview to verify feelings besides unconditional love.
Fisher et al. (2010) Rejected romantic love	Flyer, word of mouth: 'Have you just been rejected in love but can't let go?' Interview: mixed emotions obsessively thinking about beloved rejecter	10 ♀ / 5 ♂ 18-21 years old (19.8 years) Relationship: 4-48 months (21 months) End relationship: 1-32 weeks ago (9 weeks ago)	picture of beloved rejecter that stimulates feelings of intense romantic passion 30 sec	'think about events that occurred with this person'	Same instructions. Pics of: familiar acquaintance (same sex, similar age and familiarity) compared (rejecter – neutral) here to (beloved – neutral) in Aron et al (2005)	Interview to determine whether participants followed instructions and what emotions they felt during scanning
Guerra et al. (2011) Romantic and filial love	Undergraduate students in a romantic relationship reside close to family and partner have a positive relationship with father	35 ♀ 20-29 years old (21.7 years) Relationship: [unreported]	picture of boyfriend picture of father 4 sec	'view each picture for the entire time it is on screen'	Pics of: unfamiliar boyfriend of other participant unfamiliar father of other participant unfamiliar baby	Assessing valence, arousal and dominance of all pictures, using Self-Assessment Manikin (Bradley and Lang 1994)
Kim et al. (2009) Romantic love	Advertisements by broadcaster: couples who had fallen in love < 100 days ago. Were fine with appearing in a documentary on	5 ♀ / 5 ♂ 5 couples selected out of ± 100 couples 18-24 years old	5 pictures lover, taken at studio, different facial expressions. 30 sec	[not reported]	Pics of: friend (5, taken at studio) blurred human faces gray screen	

Study	Selection criteria participants	Participants	Stimuli	Instructions	Control task	Check after experiment
	love being broadcast on Korean television.	(21.1 years) Relationship length: < 100 days			Entire experiment repeated after 6 months	
Langeslag et al. (2007) Romantic love	Posters at Dutch university: in love with someone of opposite sex	9 ♀ / 9 ♂ 18-34 years old (21.5 years) Relationship: (12.1 months) 'love duration': 2.5-36 months (12.6 months)	picture beloved 0.25 sec	'focus on the fixation cross, pay attention to the faces; you'll have to answer questions about them'	Pics of: friend beautiful person (blended pics from beautiful people database) (both same sex as beloved)	At 4 random moments between pictures, a question appeared: 'was the person on the previous picture your beloved / friend/ a stranger?' to ensure participants remained alert
Langeslag, Van der Veen, Röder (2014) Romantic love	Dutch students: in love with someone of opposite sex for < 9 months	9 ♀ / 6 ♂ 18-25 years old (20.8 years) Relationship: 1-6.5 months (3.9 months) 'love duration': 2.5-8 months (5.1 months)	picture beloved (10% of trials) 0.250 sec	'respond to the target stimulus by pressing a button' Accuracy is more important than speed. [= oddball task]	Pics of: friend (10% of trials) stranger (80% of trials) (same sex as beloved, friend ± as well known)	Accuracy oddball task was 98%, which shows participants followed the instructions.
Noriuchi et al. (2008) Maternal love	Mothers of infants	13 ♀ (31.1 years old) 'age infant': (16.5 months)	Video clips of own infant, either smiling or in distress. No sound. 32 sec	[none reported]	Video clips of four unfamiliar infants, also either smiling or in distress. No sound. (wearing same clothes as own infant)	Rate 11 descriptors of feelings while viewing sample clips, selected from stimuli, on a 5-point scale. E.g. how motherly do you feel? 1-5. How much love? 1-5.
Ortigue et al. (2007) Romantic love	Advertisements to students: 'individuals who are currently intensively in love' Heterosexual Dating, engaged, or married	36 ♀ (20.1 years old) 'being in love': 1-60 months (15.3 months)	name beloved as prime 0.026 sec	'indicate as rapidly and as accurately as possible whether or not an English word was presented' [= lexical decision task	Control primes noun 'passion' (passionate hobby, e.g. science, sports, art) name 'friend'/acquaintance (same sex as beloved, similar	Check that participants were unaware of prime stimuli, through increasingly specific questions about the study.

Study	Selection criteria participants	Participants	Stimuli	Instructions	Control task	Check after experiment
	Have a favorite passion in life (e.g. science, sports, art) Report thinking about passionate hobby 60% of the day			with subliminal priming: 40 positive emotional nouns 40 pronounceable non-words 40 blank]	age and duration of friendship)	
Tiedt et al. (2014) Romantic love	in a long-term relationship (> 1 year)	14 ♀ / 14 ♂ 20-35 years old (25.5 years) Relationship: 12-96 months (40.2 months)	picture beloved 6 sec	'view the photographs and imagine a comfortable, non-erotic situation related to the person presented'	Pics: 2 friends for whom no romantic feelings (yet same sex as beloved and known > 1 year, often for longer than beloved)	Interview about experiences during experiment.
Vico et al. (2010) Love	Undergraduate students in a romantic relationship reside close to partner and four other loved ones	30 ♀ 20-27 years old	5 pictures of loved ones 4 sec / 0.5 sec	'just look at the screen and the pictures appearing in it for as long as they are on' 'try not to blink'	Pics: 5 unfamiliar people (database) 5 unfamiliar people (loved by other participants) 5 famous faces 5 unfamiliar babies	Valence, arousal and dominance of the 25 faces assessed, using the Self-Assessment Manikin (Bradley & Lang 1994)
Xu et al. (2011) Romantic love	Flyers to Beijing student email lists: "currently in a relationship and very intensely in love"	10 ♀ / 8 ♂ 19-25 years old (21.6 years) Relationship: 1.3-13 months (6.5 months)	picture beloved 30 sec	'while viewing the pictures, think about pleasurable and rewarding events (that you recalled in the pre-scan interview)'	Pics of acquaintance (same sex as beloved) 'think about neutral events (that you recalled in pre-scan interview)' Countback task	None
Zeki & Romaya (2010) Romantic love	Advertisements: 'passionately in love' In sexual relationship with lover Equal number of males and females in heterosexual and homosexual relationships	12 ♀ / 12 ♂ 19-47 years old (26.3 years) Relationship: 4 months – 23 years (3.7 years)	6-8 pictures lover 16 sec	'subjects were allowed to scan the images freely' 'to ensure consistent attention over time, participants had to press a key when a circular bulls-eye prompt appeared'	6-8 pictures of friends (same sex as lover) blank screen	Passionate Love Scale and interview to establish whether participants followed the instructions.

A study by Stoessel and colleagues was excluded as it does not address the question it purports to address: instead of comparing neural correlates in happy versus unhappy lovers (per its title), it compares neural correlates to romantic love to neural correlates to depression, without establishing whether the depressed participants also experience romantic love and thus are in fact lovers, nor establishing whether participants in romantic relationships are in fact happy (Stoessel et al. 2011).

A quick version of the same search in November 2016 resulted in two further studies. One employs the same research paradigm as most studies reported here (looking at pictures of romantic partners and friends), to find out more about the dopamine dynamics during experiences of love (Takahashi et al. 2015). The other paper reports a resting state experiment on people who are happily in love, have recently ended a relationship, or have never been in love. Researchers aim to find out whether the brain's regional homogeneity and functional connectivity differ during rest based on people's love status (Song et al. 2015). That is, in neither experiment do people have to actively perform a task and both experiments are aimed at figuring out things about the brain-during-love. These studies thus fit squarely into the analysis of CNS of love that I present in chapter 1.

Appendix B

Study	Participants	Stimuli	Instructions	Control task	Check after experiment
Chiao et al (2009)	Asian-American young adults living in Chicago, bicultural according to scales (Suinn et al. 1987). 18 ♀ / 12 ♂ 15 received individualistic prime (M age = 22.9 years old) 15 received collectivistic prime (M age = 24.8 years old)	24 personality adjectives (source not reported), in English, embedded in a sentence.	'In general, does this sentence describe you?' 'In general I am [adjective]' 'Does this sentence describe you when you are talking to your mother?' 'When talking to my mother I am [adjective]' Yes/No	'Is this sentence written in italics?' 'I am [*adjective*]' Yes/No	Whether primes functioned and participants did as expected during training. They did.
Chiao et al (2009)	24 students 12 Japanese: 5 ♀ / 7 ♂ (M age = 24 years old) 12 Caucasian Americans: 5 ♀ / 7 ♂ (M age = 21.1 years old) Singelis (1994) Self-Construal Scale to determine level of individualism/collectivism Individualism group: 7 Japanese, 3 American Collectivism group: 5 Japanese, 9 American	24 personality adjectives (source not reported), in English and Japanese (translated from the English by a bilingual translator), embedded in a sentence. 4 sec	'In general, does this sentence describe you?' 'In general I am [adjective]' 'Does this sentence describe you when you are talking to your mother?' 'When talking to my mother I am [adjective]' Yes/No	'Is this sentence written in italics?' 'I am [*adjective*]' Yes/No	
Craik, Moroz, Moscovitch et al. (1999)	4 ♀ / 4 ♂ 19-26 years old (M = 22.8 years old)	512 personality adjectives (Anderson, 1968; Kirby & Gardner, 1972), counterbalanced for valence and number of syllables 4 sec, of which 2 sec presentation time of the adjective	'How well does the adjective describe you?' [self] Almost never / rarely / sometimes / almost always	'How well does the adjective describe Brian Mulroney?' [famous other] 'How socially desirable is the trait described by this adjective?'	An unexpected Y/N recognition test (half of the adjectives only used here). Check whether time window was

Study	Participants	Stimuli	Instructions	Control task	Check after experiment
				[general semantic] Almost never / rarely / sometimes / almost always 'How many syllables does this adjective contain?' [non-semantic verbal processing, baseline] 2 / 3 / 4 / 5	comfortable; participants reported it was.
D'Argembeau et al. (2005)	7 ♀ / 6 ♂ 21-28 years old	None. Eyes closed. No motor response. 100 sec	'Now you will be asked to think about yourself for about 2 min. More specifically, I would like you to consider the traits and attitudes you have in the context of social relations. Try to think about your traits and attitudes, about your personality, in your relationship with others.' [Reflection on own traits, attitudes, personality. Three different contexts: profession, family, social relations.]	[Other-reflection:] Instructions to reflect on the personality of Jacques Chirac, singer Johnny Halliday, princess Mathilde of Belgium. [Reflection on social issues:] Instructions to reflect on social security, the right to vote, globalization of trade. [Rest:] Simply relax and open yourself to images, thoughts, memories that arise.	Post-scan verbal report of content of reflection. Ratings (on scale 1-10) of amount of images, thoughts, memories, physical sensations; of relative amount of self- and other-related thought; of relative amount of time they reflected on-topic. Participants largely did as told (1 male excluded).
D'Argembeau et al. (2007)	11 ♀ / 6 ♂ (M age = 23 years old) Prior to scanning, participants identified a close other (friend or relative).	40 personality adjectives (Klein, Loftus, and Kihlstrom 1996), translated into French. 5 sec	'You are [adjective]' 'According to [close other], you are [adjective]' Not at all / a little / quite well / completely [these are the authors' translations from the French]	'[Close other] is [adjective]' 'According to [close other], s/he is [adjective]' Not at all / a little / quite well / completely Fixation cross	

Study	Participants	Stimuli	Instructions	Control task	Check after experiment
				[baseline, 7-12 sec]	
D'Argembeau et al. (2008)	Couples of close friends or siblings (known e/o > 5 years). All were in high school 5 years ago, now in college. 10 ♀ / 10 ♂ 20-23 years old (M age = 21 years old)	40 personality adjectives (Anderson, 1968), counterbalanced for valence 3.5 sec	'At present, I am [adjective]' 'Five years ago, I was [adjective]' Yes/No	'At present, [close other] is [adjective]' 'Five years ago, [close other] was [adjective]' 'Does this adjective designate a positive trait? Yes/No	Rate the overall ease/difficulty with which each type of judgment was made
D'Argembeau et al (2010)	11 ♀ / 10 ♂ 18-28 years old (M age = 23 years old)	40 personality adjectives (Anderson, 1968), counterbalanced for valence 3.5 sec	'At present, I am [adjective]' 'Five years ago, I was [adjective]' 'In five years from now, I will be [adjective]' Yes/No	'Does this adjective designate a positive trait?' Yes/No	Rate the overall ease/difficulty with which each type of judgment was made
Fossati et al. (2003)	7 ♀ / 3 ♂ (M age = 25.8 years old)	130 personality adjectives (Anderson 1968), counterbalanced for valence 9.5 sec	'Does this trait describe you?' Yes/No	'Is this trait socially desirable?' 'Does this word contain [target letter, e.g. a] Yes/No	Surprise Y/N recognition task
Gutchess, Kensinger and Schacter (2007)	19 young (M age = 23 years old) 10 ♀ / 9 ♂ 17 elderly (M age = 72 years old) 11 ♀ / 6 ♂	288 personality adjectives (Craik et al. 1999), counterbalanced for valence 4 sec	'Make a yes/no decision for each trial and respond quickly' 'Myself (yes/no)? [adjective]' 'Einstein (yes/no)? [adjective]'	'Upper case (yes/no)? [ADJECTIVE]'	A Y/N recognition test (half of the adjectives only used here).
Heatherton et al. (2006)	16 ♀ / 14 ♂ 18-31 years old (M age = 24 years old) Prior to scanning, participants identified a best friend	270 adjectives (Anderson 1968) 2 sec	When you see SELF with an adjective, judge whether the adjective describes you. When FRIEND, judge whether the adjective describes your best friend. When CASE, judge whether the	'FRIEND [adjective]' 'CASE [adjective]' Yes/No	Surprise Y/N recognition task

Study	Participants	Stimuli	Instructions	Control task	Check after experiment
			adjective is printed in uppercase letters. 'SELF [adjective]' Yes/No		
Herwig et al. (2012)	17 ♀ / 8 ♂ 23-41 years old	Cue to indicate who to reflect on, then nothing. 9.9 sec	'The cue indicates a person who will later appear in a photograph. In the period that follows, reflect about that person, who and how she/he is or might be, for instance 'who am I, how am I as a person' or 'who and how is she/he'. When the photo-graph is presented, just look at the photo.' Reflection on self	Reflection on personal acquaintance (work/study colleague) of same gender Reflection on unknown person seen in photo prior to experiment Perception of picture of self	Non-structured interview: 2 participants excluded who reported sleepiness, 1 who moved head. Participants reported reflecting on facts about self (name, age, employment, family) and auto-biographical and future-oriented pieces of info ("I was born in.." "I want to be..", "I am..")
Jenkins and Mitchell (2011)	10 ♀ / 5 ♂ 19-32 years old (M age = 21 years old)	Three sets of 40 adjectives each, referring to personality traits, current mental states, and stable physical attributes. Lists were matched for average word length, valence, intensity. 3.6 sec	For self-trials, each question was accompanied by a chalk outline of a person's head 'In general, how [personality adjective, e.g. brave]?' 'In the moment, how [mental state adjective, e.g. bored]?' 'Physically, how [physical attribute, e.g. tall]?' 4 point scale (1 = not at all, 4 = very)	For other-trials, questions were accompanied by a picture of George W. Bush (four different pictures of him were used) 'In general, how [adjective]?' 'At the time when the photo was taken, how [adjective]?' 'Physically, how [adjective]?' 4 point scale (1 = not at all, 4 = very)	

Study	Participants	Stimuli	Instructions	Control task	Check after experiment
Johnson et al. (2002)	employees medical center 4 ♀ / 7 ♂ (M age = 35 years old)	Aurally delivered statements, via headphones. Statements that one might find on a self-report personality or mood survey: statements regarding stable traits, attitudes, abilities 4 sec: 2 sec aural presentation statement, 2 sec to respond	Make a decision about each statement. e.g. 'I get angry easily', 'my future is bright', I'd rather be alone', 'I'm good at my job', 'I can be trusted' Yes/No	Same instructions. Statements of factual knowledge, e.g. 'ten seconds is more than a minute' or 'you need water to live'. Yes/No	
Johnson et al. (2006)	Exp 1: 14 ♀ / 5 ♂ 18-27 years old (M age = 21.3 years old) Exp. 2: 8 ♀ / 3 ♂ 18-26 years old (M age = 20.8 years old) Prior to scanning, participants wrote essays on their hopes and aspirations, or on their duties and obligations.	'hopes and aspirations' or 'duties and obligations' 18 sec	'Focus on the idea expressed by the phrase and use your imagination to visualize or think about the idea… think about the idea for the entire time it is shown'	Same instructions. Phrases e.g. 'polar bears fishing' or 'pattern on oriental rug' or 'shape of USA'	
Kelley et al. (2002)	11 ♀ / 13 ♂ 18-30 years old (M age = 20 years old)	270 personality adjectives (Anderson, 1968) 2,5 sec	"Does the adjective describe you?" Yes/No	Does the adjective describe president Bush? [famous other] Is adjective presented in uppercase letters? [non-semantic] Yes/No	Surprise Y/N recognition test
Kircher et al. (2000)	Males in heterosexual relationships Generally positive attitude towards self and partner. No spectacles or facial hair on self or partner. 6 ♂ [no age info reported]	Full face, frontal color photos of self and partner. Digitized and morphed with pictures of same-sex unfamiliar other. 7 pictures drawn from endpoint self, 7 from endpoint other with which self was morphed, 7 from endpoint partner, 7 from endpoint other	Own face or unknown face? Pictures from endpoint own face. Self-descriptive adjective or not? Self-descriptive adjectives So, in 'own face' and 'self-descriptive adjectives' phases	Own face or unknown face? Pictures from endpoint unknown. Self-descriptive adjective or not? Non-self-descriptive adjectives Partner or unknown?	

Study	Participants	Stimuli	Instructions	Control task	Check after experiment
		with which partner was morphed. 232 personality adjectives (from Anderson 1968, plus an additional 32). 6 weeks prior to imaging, participants judged: 'Does this adjective describe how you typically feel and think about yourself?' (0 = extremely characteristic of me, 9 = extremely uncharacteristic). From endpoints of these ratings 8 self-descriptive adjectives and 8 non-self-descriptive adjectives were taken. 2 sec	of the experiment, self stimuli were presented one after the other, though interspersed with one (or two, for adjectives) non-self stimuli to ensure participants kept on paying attention.		
Kjaer, Nowak & Lou (2002)	2 ♀ / 5 ♂ 22-27 years old (M age = 24 years)	None. Subdued light, eye pads, earplugs. 120 sec	Reflection on own personality traits Reflection on own physical appearance	Reflection on personality traits of Danish queen Reflection on physical appearance Danish queen	Free format interview, oral report of thoughts. To ascertain instructions had been followed.
Lieberman, Jarcho, Satpute (2004)	3 ♀ / 19 ♂ 11 male soccer players; (M age = 21.7 years old) 11 improvisational actors (M age = 30.0 years old) In the end, comparison not between actors vs. athletes, but between low-experience domain vs. high-experience domain, collapsing over actors and athletes.	27 words that are unambiguously descriptive of either athletes or actors, not both, as determined in pilot study. 27 words that are not specifically associated with either actors or athletes 3 sec	'Does this word describe you? Respond as quickly and accurately as possible.' Me / not me Block of 20 athlete words, interspersed with 7 neutral words (= experimental condition for athletes) Block of 20 actor words, likewise interspersed (= experimental condition for actors)	Same instructions Block of 20 athlete words, interspersed with 7 neutral words (= control condition for actors) Block of 20 actor words, likewise interspersed (= control condition for athletes) Block of 27 neutral words	
Macrae et al. (2004)	15 ♀ / 7 ♂	540 personality adjectives (Anderson	'Does this adjective describe you?'	Fixation trials, i.e. looking at a	Surprise recognition

Study	Participants	Stimuli	Instructions	Control task	Check after experiment
	18-31 years old (M age = 20 years old)	1968), counterbalanced for word length, number of syllables, valence. 180 adjectives used per participant during scanning. 2 sec, of which 0.75 sec presentation of adjective and 1.25 sec fixation	Yes/No	cross in the middle of the screen for 2 sec An extra analysis compared neural activity during Yes responses to neural activity during No responses	test, for 180 adjectives a participant encountered during the experiment, plus 180 novel adjectives.
Modinos, Ormel, & Aleman (2009)	6 ♀ / 10 ♂ 18-27 years old (M age = 20.8 years old) Prior to scanning, participants chose an acquaintance to reflect on	Short sentences about personal qualities, attributes and attitudes 4 sec	'I am attractive' 'I often forget important things' Yes/No	'[Acquaintance] is attractive' '[acquaintance] often talks too much' 'You need water to live' 'A hand has 8 fingers' Yes/No	
Moran et al. (2006)	25 ♀ / 17 ♂ 18-33 years old (M age = 20.6 years old)	540 personality adjectives (Anderson 1968) 2 sec, of which 1.25 sec presentation of adjective and 0.75 sec fixation	judge extent to which personality characteristic is descriptive of you 1-4 (1 = not at all like me, 4 = most like me) During analysis, self-relevance was considered low for scores 1, 2 and high for scores 3, 4	Fixation cross for 2 sec Extra analysis contrasting low self-relevance to high self-relevance	
Ochsner et al. (2005)	Exp. 1: 9 ♀ / 8 ♂ (M age = 29 years old) Prior to scanning, participants identified a close other. Exp. 2: 9 ♀ / 7 ♂ (M age = 23 years old) Prior to scanning, participants identified a best friend or romantic partner [= close other] and a teaching assistant from a class taken previous semester with whom they had not	Personality adjectives (Craik et al. 1999) 2.2 sec	Exp. 1: Judge whether the adjective describes you Yes/No Exp. 2: [You about self:] Judge the extent to which the adjective describes you [You about close other] [You about non-close other]	Exp. 1: Judge whether the adjective describes [close other]. Judge whether the adjective describes a positive personality trait. Judge whether the adjective contains two syllables Yes/No Exp. 2:	

Study	Participants	Stimuli	Instructions	Control task	Check after experiment
	become friends [= non-close other]		[Close other about you:] Judge the extent to which [close other] would judge the adjective to describe you [Non-close other about you] 1-4 (1 = not at all, 4 = very much)	[Curved lines:] judge the extent to which adjective contains curved as compared to straight lines 1-4 (1 = not at all, 4 = very much)	
Schmitz, Kawahara-Baccus, & Johnson (2004)	9 ♀ / 10 ♂ (M age = 24 years old) Prior to scanning, participants identified a close friend or relative [= close other]	30 personality trait adjectives Traits: general disposition ('daring'), interactive social characteristics ('shy'), cognitive ('intelligent') and physical traits ('weak'). 3 sec	'I am [adjective]' Yes/No	'[close other] is [adjective]' 'Is the word positive?' Yes/No	
Zhu et al. (2007)	13 Chinese students: 5 ♀ / 8 ♂ (M age = 21.5 years old) 13 Western students <1year in China: 5 ♀ / 8 ♂ (M age = 23.1 years old) Groups matched on educational level and time living independently	384 personality trait adjectives, in Chinese (Liu 1990) and English (Anderson 1968) Meanings of Chinese and English adjectives were matched, valence was counterbalanced Questions in English for Westerners, in Chinese for Chinese. 3 sec, of which 2 sec for passive viewing and 1 sec for response	'Does this adjective describe you?' 'self [adjective]' during passive viewing 'self' during response Does this adjective describe your mother? Yes/No	Does this adjective describe Bill Clinton / Rongji Zhu? Is this adjective in uppercase or lowercase letters / bold- or light-faced character? Yes/No Null sessions: passively viewing rows of asterisks	Recognition memory test: old or new? 'Remember' or 'know/seems familiar'? Test included all 384 stimuli. During scanning, 192 adjectives were used.

A quick version of the same search in November 2016 resulted in 4 further studies, all of which employ a variant on the personality trait judgment task (Bergström et al. 2015; Yaoi, Osaka, and Osaka 2015; Komulainen et al. 2016; Shi et al. 2016). One study investigates whether one can tamper with the neural activity levels during self-reflection through anti-depressants (Komulainen et al. 2016). Three studies continue the quest for the neural correlates of different types / components of self-reflection and the self-reference memory effect.

Bibliography

Acevedo, Bianca P., Arthur Aron, Helen E. Fisher, and Lucy L. Brown. 2012. "Neural Correlates of Long-Term Intense Romantic Love." *Social Cognitive and Affective Neuroscience* 7 (2): 145–59. doi:10.1093/Scan/Nsq092.

Allport, Gordon W., and Henry S. Odbert. 1936. *Trait-Names: A Psycho-Lexical Study*. Vol. 47. Psychological Monographs. Albany, N.Y.: Psychological Review Company.

Anderson, Norman H. 1968. "Likableness Ratings of 555 Personality-Trait Words." *Journal of Personality and Social Psychology* 9 (3): 272-. doi:10.1037/H0025907.

Aron, Arthur, Helen E. Fisher, Debra J. Mashek, Greg Strong, Haifang Li, and Lucy L. Brown. 2005. "Reward, Motivation, and Emotion Systems Associated with Early-Stage Intense Romantic Love." *Journal of Neurophysiology* 94 (1): 327–37. doi:10.1152/jn.00838.2004.

Arpaly, Nomy. 2010. "Comment." In *Meaning in Life and Why It Matters*, 85–91. Princeton, N.J.: Princeton University Press.

Bartels, Andreas, and Semir Zeki. 2000. "The Neural Basis of Romantic Love." *Neuroreport* 11 (17): 3829–34. doi:10.1097/00001756-200011270-00046.

———. 2004. "The Neural Correlates of Maternal and Romantic Love." *Neuroimage* 21 (3): 1155–66. doi:10.1016/j.neuroimage.2003.11.003.

Beauregard, Mario, Jerome Courtemanche, Vincent Paquette, and Evelyne L. St-Pierre. 2009. "The Neural Basis of Unconditional Love." *Psychiatry Research-Neuroimaging* 172 (2): 93–98. doi:10.1016/j.pscychresns.2008.11.003.

Bechtel, William, Peter Mandik, and Jennifer Mundale. 2001. "Philosophy Meets the Neurosciences." In *Philosophy and the Neurosciences*, edited by William Bechtel, Peter Mandik, Jennifer Mundale, and Robert S. Stufflebeam, 4–22. Malden, MA: Blackwell.

Bechtel, William, and Robert S. Stufflebeam. 2001. "Epistemic Issues in Procuring Evidence about the Brain: The Importance of Research Instruments and Techniques." In *Philosophy and the Neurosciences*, edited by William Bechtel, Peter Mandik, Jennifer Mundale, and Robert S. Stufflebeam, 55–81. Malden, MA: Blackwell.

Bennett, Maxwell R., Daniel Dennett, Peter M. S. Hacker, and John Searle. 2007. *Neuroscience and Philosophy: Brain, Mind, and Language*. New York: Columbia Univ. Press.

Bennett, Maxwell R., and Peter M. S. Hacker. 2003. *Philosophical Foundations of Neuroscience*. Malden, MA: Blackwell Pub.

Bergström, Zara M., David A. Vogelsang, Roland G. Benoit, and Jon S. Simons. 2015. "Reflections of Oneself: Neurocognitive Evidence for Dissociable Forms of Self-Referential Recollection." *Cerebral Cortex* 25 (9): 2648–57. doi:10.1093/cercor/bhu063.

Berridge, Kent C., and Morten L. Kringelbach. 2008. "Affective Neuroscience of Pleasure: Reward in Humans and Animals." *Psychopharmacology* 199 (3): 457–80. doi:10.1007/s00213-008-1099-6.

Borsboom, Denny, Gideon J. Mellenbergh, and Jaap van Heerden. 2004. "The Concept of Validity." *Psychological Review* 111 (4): 1061–71. doi:10.1037/0033-295X.111.4.1061.

Buekens, Filip. 2010. "Evolutionair Revisionisme En de Integriteit van Het Manifeste Zelfbeeld." *Tijdschrift Voor Filosofie* 72 (1): 101–29.

Burton, Robert A. 2013. *A Skeptic's Guide to the Mind: What Neuroscience Can and Cannot Tell Us About Ourselves*. New York: St. Martin's Press.

Camus, Albert. 2013. *The Outsider*. Translated by Sandra Smith. London: Penguin Books.

Carskadon, Mary A., Christine Acebo, and Oskar G. Jenni. 2004. "Regulation of Adolescent Sleep: Implications for Behavior." *Annals of the New York Academy of Sciences* 1021 (1): 276–91. doi:10.1196/annals.1308.032.

Cassidy, Jude, and Phillip R Shaver. 1999. *Handbook of Attachment: Theory, Research, and Clinical Applications*. New York: Guilford Press.

Chemero, Anthony, and Charles Heyser. 2009. "Methodology and Reduction in the Behavioral Neurosciences: Object Exploration as a Case Study." In *The Oxford Handbook of Philosophy and Neuroscience*, edited by John Bickle. Oxford; New York: Oxford University Press.

Chiao, Joan Y., Tokiko Harada, Hidetsugu Komeda, Zhang Li, Yoko Mano, Daisuke Saito, Todd B. Parrish, Norihiro Sadato, and Tetsuya Iidaka. 2009a. "Dynamic Cultural Influences on Neural Representations of the Self." *Journal of Cognitive Neuroscience* 22 (1): 1–11. doi:10.1162/jocn.2009.21192.

———. 2009b. "Neural Basis of Individualistic and Collectivistic Views of Self." *Human Brain Mapping* 30 (9): 2813–20. doi:10.1002/hbm.20707.

Christoff, Kalina, Diego Cosmelli, Dorothee Legrand, and Evan Thompson. 2011. "Specifying the Self for Cognitive Neuroscience." *Trends in Cognitive Sciences* 15 (3): 104–12. doi:10.1016/j.tics.2011.01.001.

Churchland, Patricia Smith. 1986. *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. Cambridge, Mass.: MIT Press.

———. 2008. "The Impact of Neuroscience on Philosophy." *Neuron* 60 (3): 409–11. doi:10.1016/j.neuron.2008.10.023.

———. 2013. *Touching a Nerve: The Self as Brain*. New York: W.W. Norton.

Coltheart, Max. 2006a. "Perhaps Functional Neuroimaging Has Not Told Us Anything about the Mind (so Far)." *Cortex* 42 (3): 422–27.

———. 2006b. "What Has Functional Neuroimaging Told Us about the Mind (so Far)?" *Cortex* 42 (3): 323–31.

Conway, Andrew R. A., Michael J. Kane, Michael F. Bunting, D. Zach Hambrick, Oliver Wilhelm, and Randall W. Engle. 2005. "Working Memory Span Tasks: A

Methodological Review and User's Guide." *Psychonomic Bulletin & Review* 12 (5): 769–86. doi:10.3758/BF03196772.

Craik, Fergus I. M., Tara M. Moroz, Morris Moscovitch, Donald T. Stuss, Gordon Winocur, Endel Tulving, and Shitij Kapur. 1999. "In Search of the Self: A Positron Emission Tomography Study." *Psychological Science* 10 (1): 26–34. doi:10.1111/1467-9280.00102.

Crone, Eveline. 2008. *Het puberende brein: over de ontwikkeling van de hersenen in de unieke periode van de adolescentie*. Amsterdam: Bakker.

D'Argembeau, Arnaud, Dorothée Feyers, Steve Majerus, Fabienne Collette, Martial Van der Linden, Pierre Maquet, and Eric Salmon. 2008. "Self-Reflection across Time: Cortical Midline Structures Differentiate between Present and Past Selves." *Social Cognitive and Affective Neuroscience* 3 (3): 244–52. doi:10.1093/Scan/Nsn020.

D'Argembeau, Arnaud, Perrine Ruby, Fabienne Collette, Christian Degueldre, Evelyne Balteau, André Luxen, Pierre Maquet, and Eric Salmon. 2007. "Distinct Regions of the Medial Prefrontal Cortex Are Associated with Self-Referential Processing and Perspective Taking." *Journal of Cognitive Neuroscience* 19 (6): 935–44. doi:10.1162/jocn.2007.19.6.935.

D'Argembeau, Arnaud, David Stawarczyk, Steve Majerus, Fabienne Collette, Martial Van der Linden, and Eric Salmon. 2010. "Modulation of Medial Prefrontal and Inferior Parietal Cortices When Thinking about Past, Present, and Future Selves." *Social Neuroscience* 5 (2): 187–200. doi:10.1080/17470910903233562.

D'Argembeau, Arnaud, Fabienne Collette, Martial Van der Linden, Steven Laureys, Guy Del Fiore, Christian Degueldre, André Luxen, and Eric Salmon. 2005. "Self-Referential Reflective Activity and Its Relationship with Rest: A PET Study." *NeuroImage* 25 (2): 616–24. doi:10.1016/j.neuroimage.2004.11.048.

Davenport, John J, and Anthony Rudd, eds. 2001. *Kierkegaard after MacIntyre: Essays on Freedom, Narrative, and Virtue*. Chicago: Open Court.

Diamond, Lisa M., and Janna A. Dickenson. 2012. "The Neuroimaging of Love and Desire: Review and Future Directions." *Clinical Neuropsychiatry* 9 (1): 39–46.

Ebels-Duggan, Kyla. 2008. "Against Beneficence: A Normative Account of Love." *Ethics* 119 (1): 142–70.

Farah, Martha J. 2012. "Neuroethics: The Ethical, Legal, and Societal Impact of Neuroscience." *Annual Review of Psychology* 63 (1): 571–91. doi:10.1146/annurev.psych.093008.100438.

———. 2014. "Brain Images, Babies, and Bathwater: Critiquing Critiques of Functional Neuroimaging." *Hastings Center Report* 44 (s2): S19–30. doi:10.1002/hast.295.

Farah, Martha J., and Andrea S. Heberlein. 2007. "Personhood and Neuroscience: Naturalizing or Nihilating?" *Am J Bioeth* 7 (1): 37–48. doi:10.1080/15265160601064199.

Farrer, Chloe, and Chris D. Frith. 2002. "Experiencing Oneself vs Another Person as Being the Cause of an Action: The Neural Correlates of the Experience of Agency." *Neuroimage* 15 (3): 596–603. doi:10.1006/nimg.2001.1009.

Fisher, Helen E., Lucy L. Brown, Arthur Aron, Greg Strong, and Debra J. Mashek. 2010. "Reward, Addiction, and Emotion Regulation Systems Associated With Rejection in Love." *Journal of Neurophysiology* 104 (1): 51–60. doi:10.1152/jn.00784.2009.

Flanagan, Owen. 2009. "One Enchanted Being: Neuroexistentialism and Meaning." *Zygon* 44 (1): 41–49. doi:10.1111/j.1467-9744.2009.00984.x.

Fossati, Philippe, Stephanie J. Hevenor, Simon J. Graham, Cheryl Grady, Michelle L. Keightley, Fergus Craik, and Helen Mayberg. 2003. "In Search of the Emotional Self: An fMRI Study Using Positive and Negative Emotional Words." *American Journal of Psychiatry* 160 (11): 1938–45. doi:10.1176/appi.ajp.160.11.1938.

Francken, Jolien C., and Marc Slors. 2014. "From Commonsense to Science, and Back: The Use of Cognitive Concepts in Neuroscience." *Consciousness and Cognition* 29 (October): 248–58. doi:10.1016/j.concog.2014.08.019.

Frankfurt, Harry G. 1971. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68 (1): 5–20.

———. 1988. *The Importance of What We Care about: Philosophical Essays*. Cambridge, UK: Cambridge University Press.

———. 1999a. "Autonomy, Necessity, and Love." In *Necessity, Volition, and Love*, 129–41. Cambridge, UK: Cambridge University Press.

———. 1999b. *Necessity, Volition, and Love*. Cambridge, UK: Cambridge University Press.

———. 1999c. "On Caring." In *Necessity, Volition, and Love*, 155–80. Cambridge, UK: Cambridge University Press.

———. 1999d. "On the Necessity of Ideals." In *Necessity, Volition, and Love*, 108–16. Cambridge, UK: Cambridge University Press.

———. 2002. "Reply to Susan Wolf." In *Contours of Agency: Essays on Themes From Harry Frankfurt*, edited by Sarah Buss and Lee Overton, 248–249. Cambridge, MA: MIT Press.

———. 2004. *The Reasons of Love*. Princeton, N.J.: Princeton University Press.

———. 2006. *Taking Ourselves Seriously & Getting It Right*. Edited by Debra Satz. Stanford, Calif.: Stanford University Press.

Gallagher, Shaun 2003. "Phenomenology and Experimental Design. Toward a Phenomenologically Enlightened Experimental Science." *Journal of Consciousness Studies* 10 (9–10): 85–99.

———. 2013. "A Pattern Theory of Self." *Frontiers in Human Neuroscience* 7 (August). doi:10.3389/Fnhum.2013.00443.

———. 2000. "Philosophical Conceptions of the Self: Implications for Cognitive Science." *Trends in Cognitive Sciences* 4 (1): 14–21. doi:10.1016/S1364-6613(99)01417-5.

Gallagher, Shaun, and Dan Zahavi. 2008. *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science*. London: Routledge.

Gazzaniga, Michael S. 2014. *The Cognitive Neurosciences*. Cambridge, Mass.: MIT Press.

Gazzaniga, Michael S., Richard B. Ivry, and George R. Mangun. 1998. *Cognitive Neuroscience: The Biology of the Mind*. 1st ed. New York: W.W. Norton.

———. 2002. *Cognitive Neuroscience: The Biology of the Mind*. 2nd ed. New York: W.W. Norton.

———. 2008. *Cognitive Neuroscience: The Biology of the Mind*. 3rd ed. New York: W.W. Norton.

———. 2014. *Cognitive Neuroscience: The Biology of the Mind*. 4th ed. New York: W.W. Norton.

Gillihan, Seth J., and Martha J. Farah. 2005. "Is Self Special? A Critical Review of Evidence from Experimental Psychology and Cognitive Neuroscience." *Psychological Bulletin* 131 (1): 76–97. doi:10.1037/0033-2909.131.1.76.

Guerra, Pedro, Rafaela R. Campagnoli, Cynthia Vico, Eliane Volchan, Lourdes Anllo-Vento, and Jaime Vila. 2011. "Filial versus Romantic Love: Contributions from Peripheral and Central Electrophysiology." *Biological Psychology* 88 (2): 196–203.

Gusnard, Debra A., Erbil Akbudak, Gordon L. Shulman, and Marcus E. Raichle. 2001. "Medial Prefrontal Cortex and Self-Referential Mental Activity: Relation to a Default Mode of Brain Function." *Proceedings of the National Academy of Sciences* 98 (7): 4259–64. doi:10.1073/pnas.071043098.

Gutchess, Angela H., Elizabeth A. Kensinger, and Daniel L. Schacter. 2007. "Aging, Self-Referencing, and Medial Prefrontal Cortex." *Social Neuroscience* 2 (2): 117–33. doi:10.1080/17470910701399029.

Hastings, Janna, Gwen A. Frishkoff, Barry Smith, Mark Jensen, Russell A. Poldrack, Jane Lomax, Anita Bandrowski, Fahim Imam, Jessica A. Turner, and Maryann E. Martone. 2014. "Interdisciplinary Perspectives on the Development, Integration, and Application of Cognitive Ontologies." *Frontiers in Neuroinformatics* 8 (June). doi:10.3389/fninf.2014.00062.

Heatherton, Todd F., Carrie L. Wyland, C. Neil Macrae, Kathryn E. Demos, Bryan T. Denny, and William M. Kelley. 2006. "Medial Prefrontal Activity Differentiates Self from Close Others." *Social Cognitive and Affective Neuroscience* 1 (1): 18–25. doi:10.1093/scan/nsl001.

Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. "The Weirdest People in the World?" *Behavioral and Brain Sciences* 33 (2–3): 61–83. doi:10.1017/S0140525X0999152X.

Henson, Richard. 2006. "Forward Inference Using Functional Neuroimaging: Dissociations versus Associations." *Trends in Cognitive Sciences* 10 (2): 64–69. doi:10.1016/j.tics.2005.12.005.

Herwig, Uwe, Tina Kaffenberger, Caroline Schell, Lutz Jäncke, and Annette B. Brühl. 2012. "Neural Activity Associated with Self-Reflection." *BMC Neuroscience* 13 (1): 52.

Holmes, Jeremy. 1993. *John Bowlby and Attachment Theory*. London; New York: Routledge.

Husserl, Edmund. 1954. *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie. Eine Einleitung in die phänomenologische Philosophie*. Edited by Walter Biemel. Den Haag: Nijhoff.

Iacoboni, Marco. 2007. "The Quiet Revolution of Existential Neuroscience." In *Social Neuroscience : Integrating Biological and Psychological Explanations of Social*

Behavior, edited by Eddie Harmon-Jones and Piotr Winkielman. New York: Guilford Press.

"Implicit Bias & Philosophy." 2016. Accessed December 8. http://www.biasproject.org/.

Ishiguro, Kazuo. 1989. *The Remains of the Day*. New York: Knopf.

Jarrett, Christian. 2015. *Great Myths of the Brain*. Chichester: Wiley-Blackwell.

Jenkins, Adrianna C., and Jason P. Mitchell. 2011. "Medial Prefrontal Cortex Subserves Diverse Forms of Self-Reflection." *Social Neuroscience* 6 (3): 211–18. doi:10.1080/17470919.2010.507948.

Jenkins, Adrianna C., C. Neil Macrae, and Jason P. Mitchell. 2008. "Repetition Suppression of Ventromedial Prefrontal Activity during Judgments of Self and Others." *Proceedings of the National Academy of Sciences* 105 (11): 4507–12. doi:10.1073/pnas.0708785105.

Johnson, Marcia K., Carol L. Raye, Karen J. Mitchell, Sharon R. Touryan, Erich J. Greene, and Susan Nolen-Hoeksema. 2006. "Dissociating Medial Frontal and Posterior Cingulate Activity during Self-Reflection." *Social Cognitive and Affective Neuroscience* 1 (1): 56–64. doi:10.1093/scan/nsl004.

Johnson, Sterling C., Leslie C. Baxter, Lana S. Wilder, James G. Pipe, Joseph E. Heiserman, and George P. Prigatano. 2002. "Neural Correlates of Self-reflection." *Brain* 125 (8): 1808–14. doi:10.1093/brain/awf181.

Kelley, Truman L. 1927. *Interpretation of Educational Measurements*. Oxford, England: World Book Co.

Kelley, William M., C. Neil Macrae, Carrie L. Wyland, S. Caglar, S. Inati, and Todd F. Heatherton. 2002. "Finding the Self? An Event-Related fMRI Study." *Journal of Cognitive Neuroscience* 14 (5): 785–94. doi:10.1162/08989290260138672.

Kierkegaard, Søren. 1978. *Two ages: a literary review*. Edited by Howard V Hong and Edna H Hong. Princeton, N.J.: Princeton University Press.

———. 1998. *The Point of View*. Edited by Howard V Hong and Edna H Hong. Princeton, N.J.: Princeton University Press.

———. 2011. *Journals NB-NB5*. Edited by Niels J Cappelørn. Vol. 4. Kierkegaard's Journals and Notebooks. Princeton, N.J.: Princeton University Press.

Kierkegaard, Søren (Anti-Climacus). 1980. *The sickness unto death: a Christian psychological exposition for upbuilding and awakening*. Edited by Howard V Hong and Edna H Hong. Princeton, N.J.: Princeton University Press.

Kierkegaard, Søren (Johannes Climacus). 1992. *Concluding unscientific postscript to Philosophical fragments*. Edited by Howard V Hong and Edna H Hong. Princeton, N.J.: Princeton University Press.

Kim, Won, Seungyeon Kim, Jaeseung Jeong, Kyung-Uk Lee, Kook-Jin Ahn, Yong-An Chung, Keun-Young Hong, and Jeong-Ho Chae. 2009. "Temporal Changes in Functional Magnetic Resonance Imaging Activation of Heterosexual Couples for Visual Stimuli of Loved Partners." *Psychiatry Investigation* 6 (1): 19–25. doi:10.4306/pi.2009.6.1.19.

Kirby, D.M., and R.C. Gardner. 1972. "Ethnic Stereotypes: Norms on 208 Words Typically Used in Their Assessment." *Canadian Journal of Psychology/Revue Canadienne de Psychologie* 26 (2): 140–54. doi:10.1037/h0082423.

Kircher, Tilo T. J., Carl Senior, Mary L. Phillips, Philip J. Benson, Edward T. Bullmore, Mick Brammer, Andrew Simmons, Steven C. R. Williams, Mathias Bartels, and Anthony S. David. 2000. "Towards a Functional Neuroanatomy of Self Processing: Effects of Faces and Words." *Cognitive Brain Research* 10 (1–2): 133–44. doi:10.1016/S0926-6410(00)00036-7.

Kjaer, Troels W., Markus Nowak, and Hans C. Lou. 2002. "Reflective Self-Awareness and Conscious States: PET Evidence for a Common Midline Parietofrontal Core." *Neuroimage* 17 (2): 1080–86. doi:10.1006/nimg.2002.1230.

Klackl, Johannes, Eva Jonas, and Martin Kronbichler. 2012. "Existential Neuroscience: Neurophysiological Correlates of Proximal Defenses against Death-Related Thoughts." *Social Cognitive and Affective Neuroscience*, January, nss003. doi:10.1093/scan/nss003.

Klein, Stanley B., Judith Loftus, and John F. Kihlstrom. 1996. "Self-Knowledge of an Amnesic Patient: Toward a Neuropsychology of Personality and Social Psychology." *Journal of Experimental Psychology: General* 125 (3): 250.

Komulainen, Emma, Roope Heikkilä, Katarina Meskanen, Tuukka T. Raij, Lauri Nummenmaa, Jari Lahti, Pekka Jylhä, et al. 2016. "A Single Dose of Mirtazapine Attenuates Neural Responses to Self-Referential Processing." *Journal of Psychopharmacology (Oxford, England)* 30 (1): 23–32. doi:10.1177/0269881115616384.

Korsgaard, Christine M. 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford; New York: Oxford University Press.

Kroeker, Esther. 2011. "Liefde volgens Frankfurt." In *Vrijheid, noodzaak en liefde: een kritische inleiding tot de filosofie van Harry Frankfurt*, edited by Katrien Schaubroeck and Thomas Nys, translated by Katrien Schaubroeck, 139–61. Kapellen; Zoetermeer: Pelckmans; Klement.

Lamarque, Peter. 2004. "On Not Expecting Too Much from Narrative." *Mind & Language* 19 (4): 393–408.

Lammel, Stephan, Byung Kook Lim, and Robert C. Malenka. 2014. "Reward and Aversion in a Heterogeneous Midbrain Dopamine System." *Neuropharmacology* 76, Part B: 351–59. doi:10.1016/j.neuropharm.2013.03.019.

Langeslag, Sandra J. E., Bernadette M. Jansma, Ingmar H. A. Franken, and Jan W. Van Strien. 2007. "Event-Related Potential Responses to Love-Related Facial Stimuli." *Biological Psychology* 76 (1–2): 109–15. doi:10.1016/j.biopsycho.2007.06.007.

Langeslag, Sandra J. E., Frederik M. van der Veen, and Christian H. Röder. 2014. "Attention Modulates the Dorsal Striatum Response to Love Stimuli." *Human Brain Mapping* 35 (2): 503–12. doi:10.1002/hbm.22197.

Legrand, Dorothée, and Perrine Ruby. 2009. "What Is Self-Specific? Theoretical Investigation and Critical Review of Neuroimaging Results." *Psychological Review* 116 (1): 252–82. doi:10.1037/a0014172.

Leibenluft, Ellen, M. Ida Gobbini, Tara Harrison, and James V. Haxby. 2004. "Mothers' Neural Activation in Response to Pictures of Their Children and Other Children." *Biological Psychiatry* 56 (4): 225–232.

Levy, Neil. 2011. "Neuroethics: A New Way of Doing Ethics." *AJOB Neurosci* 2 (2): 3–9. doi:10.1080/21507740.2011.557683.

Libet, Benjamin, Curtis A. Gleason, Elwood W. Wright, and Dennis K. Pearl. 1983. "Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential). The Unconscious Initiation of a Freely Voluntary Act." *Brain* 106 (3): 623–42. doi:10.1093/brain/106.3.623.

Lieberman, Matthew D., Johanna M. Jarcho, and Ajay B. Satpute. 2004. "Evidence-Based and Intuition-Based Self-Knowledge: An FMRI Study." *Journal of Personality and Social Psychology* 87 (4): 421. doi:10.1037/0022-3514.87.4.421.

Lippitt, John. 2007. "Getting the Story Straight: Kierkegaard, MacIntyre and Some Problems with Narrative." *Inquiry* 50 (1): 34–69. doi:10.1080/00201740601154766.

Liu, Y. 1990. *Modern Lexicon of Chinese Frequently-Used Word Frequency*. Beijing: Space Navigation.

Lutz, Antoine, Jean-Philippe Lachaux, Jacques Martinerie, and Francisco J. Varela. 2002. "Guiding the Study of Brain Dynamics by Using First-Person Data: Synchrony Patterns Correlate with Ongoing Conscious States during a Simple Visual Task." *Proceedings of the National Academy of Sciences* 99 (3): 1586–1591. doi:10.1073/pnas.032658199.

MacIntyre, Alasdair C. 1984. *After Virtue: A Study in Moral Theory*. Notre Dame, Ind.: University of Notre Dame Press.

Macrae, C. Neil, Joseph M. Moran, Todd F. Heatherton, Jane F. Banfield, and William M. Kelley. 2004. "Medial Prefrontal Activity Predicts Memory for Self." *Cerebral Cortex* 14 (6): 647–54. doi:10.1093/cercor/bhh025.

Mashek, Debra J., Arthur Aron, and Helen E. Fisher. 2000. "Identifying, Evoking and Measuring Intense Feelings of Romantic Love." *Representative Research in Social Psychology* 24: 48–55.

Meer, Lisette van der, Sergi Costafreda, André Aleman, and Anthony S. David. 2010. "Self-Reflection and the Brain: A Theoretical Review and Meta-Analysis of Neuroimaging Studies with Implications for Schizophrenia." *Neuroscience and Biobehavioral Reviews* 34 (6): 935–46. doi:10.1016/j.neubiorev.2009.12.004.

Modinos, Gemma, Johan Ormel, and André Aleman. 2009. "Activation of Anterior Insula during Self-Reflection." Edited by Antonio Verdejo García. *PLoS ONE* 4 (2): e4618. doi:10.1371/journal.pone.0004618.

Monyer, Hannah, Frank Rösler, Gerhard Roth, Henning Scheich, Wolf Singer, Christian E. Elger, Angela D. Friederici, Christof Koch, Heiko Luhmann, and Christoph Von der Malsburg. 2004. "Das Manifest. Elf Führende Neurowissenschaftler Über Gegenwart Und Zukunft Der Hirnforschung." *Gehirn Und Geist* 6: 30–37.

Moran, Joseph M., C. Neil Macrae, Todd F. Heatherton, Carrie L. Wyland, and William M. Kelley. 2006. "Neuroanatomical Evidence for Distinct Cognitive and Affective Components of Self." *Journal of Cognitive Neuroscience* 18 (9): 1586–94. doi:10.1162/jocn.2006.18.9.1586.

Mueller, Charles W. 2004. "Conceptualization, Operationalization, and Measurement." In *The Sage Encyclopedia of Social Science Research Methods*, edited by Michael

S Lewis-Beck, Alan Bryman, and Tim Futing Liao, 162–66. Thousand Oaks, Calif.: Sage.

Murray, Ryan J., Marie Schaer, and Martin Debbané. 2012. "Degrees of Separation: A Quantitative Neuroimaging Meta-Analysis Investigating Self-Specificity and Shared Neural Activation between Self- and Other-Reflection." *Neuroscience and Biobehavioral Reviews* 36 (3): 1043–59. doi:10.1016/j.neubiorev.2011.12.013.

Nitschke, Jack B., Eric E. Nelson, Brett D. Rusch, Andrew S. Fox, Terrence R. Oakes, and Richard J. Davidson. 2004. "Orbitofrontal Cortex Tracks Positive Mood in Mothers Viewing Pictures of Their Newborn Infants." *NeuroImage* 21 (2): 583–92. doi:10.1016/j.neuroimage.2003.10.005.

Noriuchi, Madoka, Yoshiaki Kikuchi, and Atsushi Senoo. 2008. "The Functional Neuroanatomy of Maternal Love: Mother's Response to Infant's Attachment Behaviors." *Biological Psychiatry* 63 (4): 415–23. doi:10.1016/j.biopsych.2007.05.018.

Northoff, Georg, Alexander Heinzel, Moritz Greck, Felix Bermpohl, Henrik Dobrowolny, and Jaak Panksepp. 2006. "Self-Referential Processing in Our Brain - A Meta-Analysis of Imaging Studies on the Self." *Neuroimage* 31 (1): 440–57. doi:10.1016/j.neuroimage.2005.12.002.

Northoff, Georg, Pengmin Qin, and Todd E. Feinberg. 2011. "Brain Imaging of the Self - Conceptual, Anatomical and Methodological Issues." *Consciousness and Cognition* 20 (1): 52–63. doi:10.1016/j.concog.2010.09.011.

Ochsner, Kevin N., Jennifer S. Beer, Elaine R. Robertson, Jeffrey C. Cooper, John D.E. Gabrieli, John F. Kihlstrom, and Mark D'Esposito. 2005. "The Neural Correlates of Direct and Reflected Self-Knowledge." *Neuroimage* 28 (4): 797–814. doi:10.1016/j.neuroimage.2005.06.069.

O'Connor, Cliodhna, and Helene Joffe. 2013. "How Has Neuroscience Affected Lay Understandings of Personhood? A Review of the Evidence." *Public Understanding of Science* 22 (3): 254–68. doi:10.1177/0963662513476812.

Ortigue, Stephanie, Francesco Bianchi-Demicheli, A. F. D. C. Hamilton, and S. T. Grafton. 2007. "The Neural Basis of Love as a Subliminal Prime: An Event-Related Functional Magnetic Resonance Imaging Study." *Journal of Cognitive Neuroscience* 19 (7): 1218–30. doi:10.1162/jocn.2007.19.7.1218.

Ortigue, Stephanie, Francesco Bianchi-Demicheli, Nisa Patel, Chris Frum, and James W. Lewis. 2010. "Neuroimaging of Love: fMRI Meta-Analysis Evidence toward New Perspectives in Sexual Medicine." *Journal of Sexual Medicine* 7 (11): 3541–52. doi:10.1111/j.1743-6109.2010.01999.x.

Peters, Michel. 2013. "De berg van de ziel." *AdRem : Remonstrants maandblad*, April 4.

Poldrack, Russell A. 2006. "Can Cognitive Processes Be Inferred from Neuroimaging Data?" *Trends in Cognitive Sciences* 10 (2): 59–63. doi:10.1016/j.tics.2005.12.004.

Poldrack, Russell A. 2011. "Inferring Mental States from Neuroimaging Data: From Reverse Inference to Large-Scale Decoding." *Neuron* 72 (5): 692–97. doi:10.1016/j.neuron.2011.11.001.

Poldrack, Russell A., Paul C. Fletcher, Richard N. Henson, Keith J. Worsley, Matthew Brett, and Thomas E. Nichols. 2008. "Guidelines for Reporting an fMRI Study." *Neuroimage* 40 (2): 409–14. doi:10.1016/j.neuroimage.2007.11.048.

Poldrack, Russell A., Aniket Kittur, Donald Kalar, Eric Miller, Christian Seppa, Yolanda Gil, D. Stott Parker, Fred W. Sabb, and Robert M. Bilder. 2011. "The Cognitive Atlas: Toward a Knowledge Foundation for Cognitive Neuroscience." *Frontiers in Neuroinformatics* 5: 17. doi:10.3389/fninf.2011.00017.

Poldrack, Russell A., and Tal Yarkoni. 2016. "From Brain Maps to Cognitive Ontologies: Informatics and the Search for Mental Structure." *Annual Review of Psychology* 67 (1): 587–612. doi:10.1146/annurev-psych-122414-033729.

Quirin, Markus, Alexander Loktyushin, Jamie Arndt, Ekkehard Küstermann, Yin-Yueh Lo, Julius Kuhl, and Lucas Eggert. 2011. "Existential Neuroscience: A Functional Magnetic Resonance Imaging Investigation of Neural Responses to Reminders of One's Mortality." *Social Cognitive and Affective Neuroscience*, nsq106. doi:10.1093/scan/nsq106.

Ramachandran, Vilayanur S. 2011. *The Tell-Tale Brain: A Neuroscientist's Quest for What Makes Us Human*. New York: W.W. Norton.

Richards, Jessica M., Rista C. Plate, and Monique Ernst. 2013. "A Systematic Review of fMRI Reward Paradigms Used in Studies of Adolescents vs. Adults: The Impact of Task Design and Implications for Understanding Neurodevelopment." *Neuroscience & Biobehavioral Reviews* 37 (5): 976–91. doi:10.1016/j.neubiorev.2013.03.004.

Ricoeur, Paul. 1991a. "Life in Quest of Narrative." In *On Paul Ricoeur: Narrative and Interpretation*, edited by David Wood, 20–33. London; New York: Routledge.

———. 1991b. "Narrative Identity." In *On Paul Ricoeur: Narrative and Interpretation*, edited by David Wood, 188–99. London; New York: Routledge.

———. 1992. *Oneself as Another*. Chicago: University of Chicago Press.

Rogers, Timothy B., Nicholas A. Kuiper, and W.S. Kirker. 1977. "Self-Reference and the Encoding of Personal Information." *Journal of Personality and Social Psychology* 35 (9): 677–88. doi:10.1037/0022-3514.35.9.677.

Roskies, Adina. 2002. "Neuroethics for the New Millenium." *Neuron* 35 (1): 21–23.

Ruby, Perrine, and Jean Decety. 2001. "Effect of Subjective Perspective Taking during Simulation of Action: A PET Investigation of Agency." *Nature Neuroscience* 4 (5): 546–50.

Rudd, Anthony. 2007. "Kierkegaard, MacIntyre and Narrative unity—Reply to Lippitt." *Inquiry* 50 (5): 541–549. doi:10.1080/00201740701612416.

———. 2012. *Self, Value, and Narrative: A Kierkegaardian Approach*. Oxford: Oxford University Press.

Rudd, Anthony, and John J. Davenport, eds. 2015. *Love, Reason, and Will: Kierkegaard after Frankfurt*. New York: Bloomsbury.

Satel, Sally L., and Scott O. Lilienfeld. 2013. *Brainwashed: The Seductive Appeal of Mindless Neuroscience*. New York: Basic Books.

Schechtman, Marya. 1990. "Personhood and Personal Identity." *Journal of Philosophy* 87 (2): 71–92. doi:10.2307/2026882.

———. 1996. *The Constitution of Selves*. Ithaca, NY: Cornell University Press.
———. 2007. "Stories, Lives, and Basic Survival: A Refinement and Defense of the Narrative View." In *Narrative and Understanding Persons*, edited by Daniel D. Hutto, 155–78. Cambridge: Cambridge University Press.
———. 2011. "The Narrative Self." In *The Oxford Handbook of the Self*, edited by Shaun Gallagher, 394–416. Oxford Handbooks in Philosophy. Oxford; New York: Oxford University Press.
Schmitz, Taylor W., Tisha N. Kawahara-Baccus, and Sterling C. Johnson. 2004. "Metacognitive Evaluation, Self-Relevance, and the Right Prefrontal Cortex." *NeuroImage* 22 (2): 941–47. doi:10.1016/j.neuroimage.2004.02.018.
Sellars, Wilfrid. 1962. "Philosophy and the Scientific Image of Man." In *Frontiers of Science and Philosophy*, edited by Robert Colodny, 35–78. Pittsburgh: University of Pittsburgh Press.
Senior, Carl, Tamara Russell, and Michael S Gazzaniga. 2009. *Methods in Mind*. Cambridge, Mass.: MIT Press.
Shi, Zhenhao, Yina Ma, Bing Wu, Xinhuai Wu, Yuanye Wang, and Shihui Han. 2016. "Neural Correlates of Reflection on Actual versus Ideal Self-Discrepancy." *NeuroImage* 124, Part A (January): 573–80. doi:10.1016/j.neuroimage.2015.08.077.
Singelis, Theodore M. 1994. "The Measurement of Independent and Interdependent Self-Construals." *Personality and Social Psychology Bulletin* 20 (5): 580–591.
Smith, Zadie. 2012. *NW*. New York: Penguin Press.
Song, Hongwen, Zhiling Zou, Juan Kou, Yang Liu, Lizhuang Yang, Anna Zilverstand, Federico d'Oleire Uquillas, and Xiaochu Zhang. 2015. "Love-Related Changes in the Brain: A Resting-State Functional Magnetic Resonance Imaging Study." *Frontiers in Human Neuroscience* 9 (February). doi:10.3389/fnhum.2015.00071.
Stee, Annemarie van. forthcoming. "Cognitive Neuroscience and Self-Understanding." In *What Can the Humanities Contribute to Our Practical Self-Understanding?*, edited by Marcus Düwell.
———. 2015a. "Liefde En Persoonlijke Identiteit: Alledaagse Liefde, Ambivalente Identiteit." *Algemeen Nederlands Tijdschrift Voor Wijsbegeerte* 107 (1): 49–70. doi:10.5117/ANTW2015.1.STEE.
———. 2015b. "Selves, Existentially Speaking." In *Love, Reason, and Will: Kierkegaard after Frankfurt*, edited by Anthony Rudd and John J Davenport, 177–95. New York: Bloomsbury.
Stoessel, Christina, Juliane Stiller, Stefan Bleich, Dominikus Boensch, Arnd Doerfler, Meritxell Garcia, Tanja Richter-Schmidinger, Johannes Kornhuber, and Clemens Forster. 2011. "Differences and Similarities on Neuronal Activities of People Being Happily and Unhappily in Love: A Functional Magnetic Resonance Imaging Study." *Neuropsychobiology* 64 (1): 52–60. doi:10.1159/000325076.
Stokes, Patrick. 2010. "Naked Subjectivity: Minimal vs. Narrative Selves in Kierkegaard." *Inquiry* 53 (4): 356–82. doi:10.1080/00201741x.2010.493370.
Strawson, Peter F. 1974. *Freedom and Resentment, and Other Essays*. London: Methuen.

Suinn, Richard M., Kathryn Rickard-Figueroa, Sandra Lew, and Patricia Vigil. 1987. "The Suinn-Lew Asian Self-Identity Acculturation Scale: An Initial Report." *Educational and Psychological Measurement* 47 (2): 401–407.

Takahashi, Kayo, Kei Mizuno, Akihiro T. Sasaki, Yasuhiro Wada, Masaaki Tanaka, Akira Ishii, Kanako Tajima, et al. 2015. "Imaging the Passionate Stage of Romantic Love by Dopamine Dynamics." *Frontiers in Human Neuroscience* 9 (April). doi:10.3389/fnhum.2015.00191.

Tallis, Raymond. 2011. *Aping Mankind: Neuromania, Darwinitis and the Misrepresentation of Humanity*. Durham, UK: Acumen Pub.

Taylor, Charles. 1971. "Interpretation and Sciences of Man." *Review of Metaphysics* 25 (1): 3–51.

———. 1985a. "Cognitive Psychology." In *Philosophical Papers Vol. 1: Human Agency and Language*, 187–212. Cambridge, UK: Cambridge University Press.

———. 1985b. *Philosophical Papers Vol. 1: Human Agency and Language*. Cambridge, UK: Cambridge University Press.

———. 1985c. "Self-Interpreting Animals." In *Philosophical Papers Vol. 1: Human Agency and Language*, 45–76. Cambridge, UK: Cambridge University Press.

———. 1989. *Sources of the Self : The Making of the Modern Identity*. Cambridge, Mass.: Harvard University Press.

Tiedt, Hannes O., Klaus M. Beier, Andreas Lueschow, Alfred Pauls, and Joachim E. Weber. 2014. "A Different Pattern of Lateralised Brain Activity during Processing of Loved Faces in Men and Women: A MEG Study." *Biological Psychology* 103 (December): 255–61. doi:10.1016/j.biopsycho.2014.09.014.

Tretter, Felix, and Boris Kotchoubey. 2014. "Memorandum „Reflexive Neurowissenschaft"." http://www.psychologie-heute.de/home/lesenswert/memorandum-reflexive-neurowissenschaft/.

Uttal, William R. 2001. *The New Phrenology: The Limits of Localizing Cognitive Processes in the Brain*. Cambridge, Mass.: MIT Press.

Varela, Francisco J. 1996. "Neurophenomenology: A Methodological Remedy for the Hard Problem." *Journal of Consciousness Studies* 3 (4): 330–49.

Vico, Cynthia, Pedro Guerra, Humbelina Robles, Jaime Vila, and Lourdes Anllo-Vento. 2010. "Affective Processing of Loved Faces: Contributions from Peripheral and Central Electrophysiology." *Neuropsychologia* 48 (10): 2894–2902.

Vogeley, Kai, and Shaun Gallagher. 2011. "Self in the Brain." In *The Oxford Handbook of the Self*, edited by Shaun Gallagher, 111–36. Oxford: Oxford University Press.

Whitfield-Gabrieli, Susan, Joseph M. Moran, Alfonso Nieto-Castañón, Christina Triantafyllou, Rebecca Saxe, and John D. E. Gabrieli. 2011. "Associations and Dissociations between Default and Self-Reference Networks in the Human Brain." *NeuroImage* 55 (1): 225–32. doi:10.1016/j.neuroimage.2010.11.048.

Wolf, Susan R. 1997. "Meaningful Lives in a Meaningless World." *Quaestiones Infinitae* 19: 1–22.

———. 2002. "The True, the Good, and the Lovable: Frankfurt's Avoidance of Objectivity." In *Contours of Agency Essays on Themes from Harry Frankfurt*, edited by Sarah Buss and Lee Overton, 227–44. Cambridge, MA: MIT Press.

———. 2010. *Meaning in Life and Why It Matters*. Princeton, N.J.: Princeton University Press.

Xu, Xiaomeng, Arthur Aron, Lucy L. Brown, Guikang Cao, Tingyong Feng, and Xuchu Weng. 2011. "Reward and Motivation Systems: A Brain Mapping Study of Early-Stage Intense Romantic Love in Chinese Participants." *Human Brain Mapping* 32 (2): 249–57. doi:10.1002/Hbm.21017.

Yaoi, Ken, Mariko Osaka, and Naoyuki Osaka. 2015. "Neural Correlates of the Self-Reference Effect: Evidence from Evaluation and Recognition Processes." *Frontiers in Human Neuroscience* 9: 383. doi:10.3389/fnhum.2015.00383.

Young, Larry J. 2009. "Being Human: Love: Neuroscience Reveals All." *Nature* 457 (7226): 148. doi:10.1038/457148a.

Zahavi, Dan. 2003. *Husserl's Phenomenology*. Cultural Memory in the Present. Stanford, CA: Stanford University Press.

———. 2005. *Subjectivity and Selfhood: Investigating the First-Person Perspective*. Cambridge, Mass.: MIT Press.

Zahavi, Dan, and Andreas Roepstorff. 2011. "Faces and Ascriptions: Mapping Measures of the Self." *Conscious Cogn* 20 (1): 141–48. doi:10.1016/j.concog.2010.10.011.

Zeki, Semir, and John Paul Romaya. 2010. "The Brain Reaction to Viewing Faces of Opposite- and Same-Sex Romantic Partners." *PLoS ONE* 5 (12): e15802. doi:10.1371/journal.pone.0015802.

Zhu, Ying, Li Zhang, Jin Fan, and Shihui Han. 2007. "Neural Basis of Cultural Influence on Self-Representation." *NeuroImage* 34 (3): 1310–16. doi:10.1016/j.neuroimage.2006.08.047.

Samenvatting

Inzicht in existentieel zelfinzicht

Een analyse van bijdragen uit cognitieve neurowetenschap en filosofie

Anna voelt zich vervreemd van haar man en vraagt zich af of ze bij hem zal blijven of niet. Clemens heeft zijn vrouw en drie kinderen verloren bij een auto-ongeluk. Hij vraagt zich af wie hij nu nog is. 'Wie ben ik nou eigenlijk? Wat doet er werkelijk toe voor mij?' Wanneer mensen op zoek zijn naar zelfinzicht, zijn ze geregeld op zoek naar inzicht in wat voor hen van doorslaggevend belang is. De mensen van wie ze houden, bijvoorbeeld, of ook projecten, al dan niet professioneel, waar ze hun identiteit mee verbonden hebben.

Ook wanneer mensen niet expliciet reflecteren op wie ze zijn, maar gewoon opgaan in het leven van alledag, geeft wat voor hen van fundamenteel belang is automatisch richting aan hun leven. Mensen zijn vanzelf gemotiveerd om actie te ondernemen ten behoeve van hun belangrijkste mensen en projecten. Bovendien beïnvloedt het reilen en zeilen van hun naasten hoe het met hen gaat. Edward bijvoorbeeld is butler en druk bezig met zijn werk. Hij realiseert zich niet dat hij houdt van het hoofd van de huishouding, Ms. Kenton, maar hij is blij in haar nabijheid en wanneer hij ziet dat zij emotioneel is, ontroert hem dat ook hevig.

Ik spreek over het existentiële zelf om te refereren naar dit aspect van het menselijk zelf: het aspect dat gevormd wordt door wat voor iemand van doorslaggevend belang is. Wanneer mensen reflecteren op het existentiële aspect van hun leven vormen ze een begrip van dat aspect aan wie ze zijn, wat ik existentieel zelfbegrip noem. De levens van mensen worden mede gevormd door wat er voor hen werkelijk toe doet en door het begrip dat ze zich daarvan vormen. Mijn doel is dit existentiële aspect van het menselijk bestaan beter te begrijpen.

Het existentiële aspect van het menselijke bestaan wordt traditioneel door filosofen en andere geesteswetenschappers onderzocht. Tegenwoordig is het echter ook mogelijk om cognitief neurowetenschappelijk onderzoek te doen naar, bijvoorbeeld, liefde en zelfreflectie. Cognitieve neurowetenschappers onderzoeken de hersenprocessen die ten grondslag liggen aan mentale processen. Sommigen denken dat

dergelijk onderzoek ons begrip van het menselijke bestaan danig zal vergroten. Experimenteel onderzoek zou zekerder zijn dan filosofisch onderzoek, zo redeneert men. Filosofen baseren hun inzichten op intuïties en interpretaties, terwijl neurowetenschappers experimenten doen. Anderen beweren dat cognitieve neurowetenschap weinig tot geen interessante inzichten te bieden heeft. Het zou nooit filosofische analyses kunnen vervangen. De discussie is vaak polemisch. Ik beoog systematisch te onderzoeken wat voor typen inzicht we kunnen verwachten van cognitieve neurowetenschap en van filosofie. Ik beschouw wat de status is van die bijdragen in het licht van de methoden waarmee ze verkregen zijn. Wat dragen filosofie en cognitieve neurowetenschap (CNS) bij aan ons begrip van het existentiële aspect aan het menselijk bestaan?

Kortgezegd blijkt uit mijn analyse dat CNS inzicht bijdraagt in de neurale mechanismen die ons in staat stellen te zijn wie we zijn en te doen wat we doen. Filosofie op het gebied van de menselijke existentie draagt bij aan zelfinzicht door structurele kenmerken van onze geleefde ervaring te verduidelijken en de vragen te verhelderen die we onszelf kunnen stellen wanneer we zoeken naar zelfinzicht. De bijdragen van CNS en filosofie aan zelfinzicht zijn heel verschillend en de twee disciplines zijn dus niet in competitie met elkaar. Filosofie berust op interpretatie, maar gaat daar in principe systematisch mee om, door voortdurend aannames op te graven die ten grondslag liggen aan interpretaties, systematisch op die aannames te reflecteren en te beargumenteren waarom ze wel of niet gegrond zijn. In het neurowetenschappelijke onderzoek zitten momenten van interpretatie besloten waar op dit moment weinig of niet op wordt gereflecteerd. Dit is een van de belangrijkste punten waarop neurowetenschap en filosofie zouden kunnen samenwerken.

Cognitieve neurowetenschap en interpretatie

Cognitieve neurowetenschappers onderzoeken de neurale correlaten van mentale processen. In de eerste twee hoofdstukken analyseer ik het bestaande cognitief neurowetenschappelijke onderzoek naar liefde en naar zelfreflectie. Uit mijn analyse blijkt dat cruciale punten in CNS onderzoek berusten op interpretatieve keuzes. Allereerst is er de manier waarop het onderzoek wordt opgezet. Selectiecriteria voor deelname en vooral de taken die deelnemers moeten verrichten terwijl hun hersenactiviteit gemeten wordt, zorgen ervoor dat ingezoomd wordt op een deelaspect van liefde of zelfreflectie. Zo worden deelnemers aan het CNS onderzoek naar romantische liefde geworven met behulp van de vraag: ‘Are you truly madly deeply in love?’ (‘Ben jij hoteldebotel tot over je oren verliefd?’). Terwijl hun hersenactiviteit gemeten wordt, kijken deelnemers naar foto’s van hun geliefde zonder verder iets te

doen. Zo is het CNS onderzoek naar romantische liefde in de praktijk onderzoek naar de neurale correlaten van de ervaringen die tomeloos verliefde mensen hebben wanneer ze een foto zien van hun geliefde. Ten tweede interpreteren onderzoekers wat de gevonden neurale resultaten zouden kunnen betekenen ten aanzien van het mentale proces dat werd onderzocht. Ze suggereren dat de neurale activiteit die ze in hun onderzoek hebben gevonden hun idee onderbouwt dat liefde vooral een kwestie van motivatie is, meer dan van emotie. Of dat de gevonden neurale activiteit het neurale mechanisme zou kunnen zijn waardoor liefde blind maakt, of waardoor liefde verslavend is.

Beide momenten van interpretatie zijn onoverkomelijk: je hebt nou eenmaal een taak nodig om CNS onderzoek te kunnen doen en aangezien CNS beoogt de verbanden tussen neurale en mentale processen bloot te leggen, zullen er altijd hypotheses moeten worden opgesteld over wat neurale resultaten te zeggen zouden kunnen hebben over het mentale proces dat onderzocht werd. Er zijn echter goede en minder goede manieren om met deze cruciale momenten van interpretatie om te gaan. Allereerst is het belangrijk dat onderzoekers zich bewust zijn van de grenzen van het onderzoek dat ze doen en die grenzen ook communiceren. Het onderzoek dat ik net beschreef, gaat niet over romantische liefde in het algemeen, laat staan over liefde in het algemeen, waar behalve gevoelens bijvoorbeeld ook gedragingen en overtuigingen bij komen kijken. Bovendien is niet alle romantische liefde van het type 'hoteldebotel verliefd'. En het idee dat liefde verslavend is, is geen onderzoeksbevinding, maar een interpretatie van de onderzoeksresultaten waar veel haken en ogen aan zitten. Er zijn allerlei andere interpretaties mogelijk. Omdat interpretaties vaak herhaald worden en ook omdat ze niet altijd als interpretaties herkend worden maar voor onderzoeksresultaten worden aangezien, gaat het op dit punt vaak mis. Zo kan neurowetenschappelijk onderzoek ongemerkt bestaande ideeën over liefde retorische kracht verlenen in plaats van ze te testen.

Cognitieve neurowetenschappers reflecteren grotendeels niet op deze momenten van interpretatie in hun onderzoek, niet in tekstboeken, handboeken, wetenschappelijke tijdschriften en niet op blogs. Ze worden hier niet voor opgeleid. Bovendien zijn er al zoveel andere factoren in het onderzoeksproces die kritische reflectie behoeven. Met name de statistische analyses die toegepast kunnen worden op onderzoeksdata zijn recent veel bediscussieerd. Een belangrijk middel om problemen daarmee aan te pakken bestaat uit het opzetten van review artikelen en databases waarin de resultaten van zo veel mogelijk studies worden vergeleken.

Ook voor zulke meta-analyses toon ik aan hoe belangrijk het is kritisch te reflecteren op interpretaties, zowel qua taken als qua interpretatie van resultaten. Op dit moment worden in review artikelen en databases ongemerkt studies samengenomen

die op verschillende aspecten van een mentaal proces inzoomen. Zo blijkt uit mijn analyse dat er in het CNS onderzoek naar zelfreflectie met minstens negen verschillende labels (bv. *self-referential processing*, *self-reflection*, *self-awareness*) naar twee verschillende taken wordt gerefereerd. Bij de ene taak krijgen mensen de tijd om vrijuit op hun persoonlijkheid te reflecteren, bij de andere moeten ze binnen een paar seconden beslissen of een persoonlijkheidskenmerk op hen van toepassing is of niet. Het is goed mogelijk dat verschillende mentale processen bij deze twee taken betrokken zijn. Het verschilt per review artikel welke studies met welke labels wel of niet worden meegenomen. Andersom gebeurt het ook dat onderzoeken die dezelfde taak gebruiken en waarbij dus dezelfde mentale processen een rol spelen toch niet met elkaar vergeleken worden. In het CNS onderzoek naar moederliefde wordt dezelfde taak gebruikt als in het CNS onderzoek naar hechting tussen moeders en kinderen, maar doordat er een ander label op wordt geplakt, komen ze niet in dezelfde meta-analyses terecht.

In hoofdstuk 4 introduceer ik een conceptuele vorm van review als oplossing voor deze problemen. Waar hoofdstukken 1 en 2 al toonden wat een conceptuele review is en wat zo'n review kan doen, maak ik dit in hoofdstuk 4 expliciet. Op basis van een inventaris van de keuzes die gemaakt worden bij het opzetten van onderzoek, worden in een conceptuele review de conceptuele aannames omtrent het onderwerp van onderzoek naar boven gehaald en kritisch bevraagd. Aangezien het opgraven en bevragen van aannames het dagelijks werk is van filosofen, is dit een punt waarop neurowetenschappers en filosofen zouden kunnen samenwerken. Het is dan wel cruciaal dat de filosofen in kwestie goed op de hoogte zijn van de kaders waarbinnen neurowetenschappers moeten werken. Het zijn vaak die kaders die maken dat neurowetenschappers hun taken opzetten op de manieren die ze doen. Die taken hebben vervolgens wel conceptuele consequenties. Een conceptuele review verheldert welke consequenties dat zijn. Dit helpt om de juiste conclusies te communiceren en het helpt bij het opzetten van overzichtsartikelen en databases van kwaliteit. Door te verduidelijken op welk aspect van een proces er wordt ingezoomd, wordt bovendien duidelijk waar lacunes liggen en dus mogelijkheden voor verder onderzoek. Ook dat is voor neurowetenschappers een voordeel.

Cognitieve neurowetenschap en zelfinzicht

Op dit moment draagt neurowetenschappelijk onderzoek naar liefde en naar zelfreflectie nog weinig tot niets bij aan ons begrip van existentieel zelfinzicht. De conclusies over de neurale activiteit zijn simpelweg nog veel te voorlopig. Maar hoe zit

dit met bijdragen van CNS aan zelfinzicht in bredere zin? En hoe zal zich dit in de toekomst ontwikkelen?

In het laatste deel van hoofdstuk 4 beargumenteer ik dat de belangrijkste bijdrage die CNS te leveren heeft aan zelfinzicht bestaat uit het leveren van inzicht in de neurale mechanismen die ons in staat stellen te doen wat we doen en te ervaren wat we ervaren. Dit is vooral relevant voor zelfinzicht wanneer er iets structureel misgaat doordat de neurale mechanismen haperen. Wanneer we niet in staat zijn ook maar ergens blij over te worden, laat staan liefde te ervaren, dan interesseert het ons dat dit te maken zou kunnen hebben met een gebrekkig functionerend dopamine-systeem in het midden van het brein. Verder onderstreept neurowetenschappelijk onderzoek de rol die automatismen spelen bij de totstandkoming van gedrag, zoals psychologisch onderzoek ook (en op directere wijze) aantoont. Het kan een bijdrage leveren aan filosofie waar filosofische visies iets poneren over het brein en de betrokkenheid van het brein bij mentale processen.

Tegelijkertijd kent de relevantie van deze bijdragen zijn beperkingen. Wanneer we op zoek zijn naar zelfinzicht, bijvoorbeeld met betrekking tot waar we van houden, dan zijn we doorgaans niet op zoek naar de informatie dat onze liefdeservaringen gepaard gaan met activiteit in het dopamine-systeem in het midden van het brein. We zoeken niet naar inzicht in de onderliggende neurale mechanismen van onze geleefde ervaring, maar naar een beter begrip van die geleefde ervaring zelf.

Filosofie, zelfinzicht en interpretatie

In hoofdstuk 3 bespreek ik meerdere filosofen die over het existentiële aspect van het menselijk bestaan hebben nagedacht. Zo bespreek ik Susan Wolfs analyses van wat een leven zinvol maakt; Harry Frankfurts ideeën over het verband tussen waar wij van houden en wie wij zijn; Charles Taylors visie over hoe ons zelfbegrip vormend is voor wie wij zijn; en Søren Kierkegaards beschouwing van hoe onze zelfverhouding bepalend is voor wie we aan het worden zijn. Telkens evalueer ik in hoeverre zij bijdragen aan ons inzicht in situaties als die van Anna, Clemens en Edward. Aan het eind van hoofdstuk 3 reflecteer ik expliciet op de typen bijdragen die de filosofen leveren en op de manier waarop ze omgaan met interpretatie.

Allereerst verhelderen deze filosofen allerlei geleefde ervaringen, door de structurele kenmerken van deze ervaringen precies te beschrijven evenals de structurele verbanden tussen verschillende ervaringen. Harry Frankfurts analyses over liefde en het verband tussen liefde en de stabiele tendensen in ons doen en laten—onze praktische identiteit—zijn hier een voorbeeld van. Waar we van houden, laat ons niet koud. We zijn automatisch gemotiveerd om op zo'n manier te handelen dat we bijdragen aan het

welzijn van onze geliefden. Bovendien zijn we in het algemeen niet in staat om te bepalen van wie we houden; we zijn nu eenmaal gehecht aan mensen als onze ouders, broers en zussen, een liefdespartner of een vriend voor het leven. Op deze manieren maakt liefde ons onvrij; het bindt onze wil. Tegelijkertijd werkt liefde bevrijdend. Ook dit kan begrepen worden door in ogenschouw te nemen dat we automatisch gemotiveerd zijn om te handelen ten bate van onze geliefden en dat onze geliefden ons niet koud laten. Het is makkelijker om keuzes te maken wanneer we iets in ons leven hebben dat van ultiem belang voor ons is. Op die manier werkt liefde bevrijdend. En wanneer we handelen uit liefde, ervaren we dit als vrij handelen, als handelen waar we achter staan. Liefde bindt ons en liefde bevrijdt ons. Dat lijkt een tegenspraak, maar wanneer we het structurele verband tussen liefde en onze praktische identiteit precies beschouwen, begrijpen we hoe we beide tegelijkertijd kunnen ervaren.

Ten tweede verhelderen de filosofen bestaande manieren van denken over geleefde ervaring en nemen die kritisch onder de loep. Ze argumenteren voor de geldigheid van sommige aannames en de ongeldigheid van andere. Bij het uitdiepen van het structurele verband tussen liefde en praktische identiteit stelt Frankfurt dat wij de belangen van geliefden omarmen als de onze. Dit is een zaak van de wil, volgens Frankfurt, en het kan los staan van onze precieze gevoelens voor onze geliefden. Ik verduidelijk dat deze analyse door Frankfurt begrepen moet worden in het licht van zijn ambitie om andere filosofen ervan te overtuigen dat liefde bron kan zijn van geldige redenen. Ook legt Frankfurt zo een verband met zijn eerdere werk over wilsvrijheid. Zo verhelder ik de context waarbinnen Frankfurt zijn analyse ontwikkelt, maar vervolgens reflecteer ik er ook kritisch op. Frankfurts analyse kan bijvoorbeeld geen recht doen aan de kleine of grotere ambivalenties waarmee veel liefdesrelaties gepaard gaan. Een broer kan heimelijk verlangen dat zijn zus geen promotie krijgt, omdat hij er niet goed tegen kan dat zij succesvoller is dan hij. Ook al is hij onherroepelijk met zijn zus verbonden, hij omarmt haar belangen niet volledig als de zijne en deze ambivalenties in de wil staan niet los van zijn ambivalente gevoelens. Het relatiemodel dat ik tegenover Frankfurts identificatiemodel zet, kan beter recht doen aan deze ambivalenties. Ik betoog in hoofdstuk 5 dat het ook op andere manieren een betere analyse geeft van het existentiële aspect van het menselijk bestaan.

Ten derde ontwikkelen veel existentiefilosofen een visie op het goede leven, of in andere woorden, een normatieve visie op de menselijke existentie. Susan Wolf denkt na over wat een leven zinvol maakt. Haar uiteenzetting van de criteria waaraan een leven moet voldoen wil het een zinvol leven zijn is zowel een verhelderende analyse van criteria die vandaag de dag impliciet gehanteerd worden alsook een argument over wat in het leven het nastreven waard is. Door heel hoofdstuk 3 heen evalueer ik op welke manieren existentiefilosofen als Wolf, Frankfurt, Taylor en Kierkegaard wel of niet

bijdragen aan ons inzicht in het existentiële aspect van het menselijk bestaan. Het blijkt dat met name hun normatieve ambities geregeld in de weg zitten van een daadwerkelijke verheldering van de existentiële fenomenen.

In algemenere zin blijkt dat filosofie in hoge mate een kwestie is van interpretatie. Filosofen proberen hier zo systematisch en zelfkritisch mogelijk mee om te gaan. Het systeem is nooit helemaal waterdicht. Fenomenen kunnen over het hoofd gezien worden. Het belang van alles wat zich aan de reflectie onttrekt, kan gemakkelijk onderschat worden, zoals bijvoorbeeld het belang van automatismen voor ons doen en laten. Maar systemisch bevordert de academische filosofie zelfkritische reflectie, door het opgraven en kritisch bevragen van aannames onder gangbare manieren van denken, inclusief de eigen manieren van denken.

Existentieel zelfinzicht

In hoofdstuk 5 zet ik uiteen wat we op basis van de eerdere hoofdstukken te weten zijn gekomen over existentieel zelfinzicht. Ik bouw daarbij met name voort op de besproken filosofen door een synthese te vormen van hun uiteenlopende ideeën en die te verbeteren en aan te vullen waar nodig. Zo ontstaat een relationele visie op het existentiële zelf en existentieel zelfbegrip. Het existentiële aspect van het zelf wordt gevormd door de relaties waarin mensen staan tot diegenen en datgene wat voor hen van fundamenteel persoonlijk belang is. Deze relaties zijn affectief-volitioneel, wat zoveel wil zeggen als dat het voelen en willen van mensen onlosmakelijk met elkaar verbonden zijn. In hoofdstuk 5 leg ik precies uit wat dit betekent. Ik beargumenteer dat de relationele visie te verkiezen is boven Frankfurts visie omdat hij meer recht doet aan ambivalenties in persoonlijke relaties en aan de intersubjectiviteit van die relaties. Ook zet de relationele visie een praktisch probleem op de kaart: hoe moeten we omgaan met de mensen die van fundamenteel persoonlijk belang voor ons zijn?

Tussen wie wij zijn en wie wij denken te zijn, vindt een constant heen en weer plaats. Wanneer we onszelf verrassen, doordat we ons anders gedragen of voelen in een bepaalde situatie dan we hadden verwacht op basis van ons zelfbegrip, dan is dat een reden om op zoek te gaan naar zelfinzicht. Soms moeten we dan concluderen dat we niet genoeg rekening hebben gehouden met ons brein, wanneer we bijvoorbeeld al tijden niet genoeg hebben geslapen. Er is dan niet aan neurale instaat stellende voorwaarden voldaan. Doorgaans zoeken we echter naar redenen voor patronen in ons handelen en ervaren en vinden die op het niveau van de geleefde ervaring zelf in plaats van op een onderliggend neuraal niveau. Edward ontdekt misschien dat Ms. Kenton veel belangrijker voor hem is dan hij altijd in de gaten had en Anna ontwikkelt een nieuwe visie op de relatie tot haar man. Het zelfbegrip dat zich zo ontwikkelt, heeft gevolgen

voor toekomstige handelingen en ervaringen. Edward zal Ms. Kenton anders benaderen als hij aan zichzelf toegeeft hoeveel hij om haar geeft. Anna overweegt om bij haar man weg te gaan en of ze dit ook daadwerkelijk doet of niet, hangt af van de visie op de relatie die ze ontwikkelt. Wie mensen zijn, is voor een groot deel niet aan hen—Edward kan er niet zoveel aan doen dat hij houdt van Ms. Kenton. Maar juist het vermogen een begrip te vormen van wie zij zijn en daarnaar te handelen, maakt dat mensen toch ook richting geven aan hun leven. Zo beïnvloeden wij mensen wie wij worden en wie wij, al wordende, zijn.

Curriculum Vitae

Annemarie van Stee (Meppel, 1983) is a philosopher who likes to submit existential issues to clear-headed philosophical analysis. This places her research interests at the intersection of philosophical anthropology and ethics. She has published articles on the relation between love and selfhood; on an existential take on religious faith; and on what cognitive neuroscience may contribute to self-understanding.

Van Stee obtained a bachelor's degree at University College Utrecht (*cum laude*), an MSc degree in Cognitive Neuroscience (CNS) (2008, *cum laude*) and an MA degree in Philosophy (2011, *cum laude*), both at Radboud University Nijmegen. Before studying, she spent a year in Norway and during her bachelor, a semester at the National University of Singapore. Her research master in CNS included a yearlong traineeship at the Max Planck Institute for Psycholinguistics. Van Stee's MA-thesis 'Philosophical anthropological assumptions in cognitive neuroscience of self' received the ESSSAT Student Prize 2012.

Van Stee conducted her PhD research within the NWO program 'What can the humanities contribute to our practical self-understanding?' and was based at Leiden University. She spent the Spring semester of 2014 as a visiting researcher at the Center for Subjectivity Research in Copenhagen. From 2014 to 2016, she was also affiliated with the Department of Philosophy and Religious Studies at Utrecht University.

Van Stee currently works at the Faculty of Philosophy, Theology and Religious Studies at Radboud University in Nijmegen, where she prepares a research proposal on assumptions about meaning and meaninglessness in relation to the Dutch 'completed life' ('voltooid leven') debate.