Consciousness and Cognition xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Faces and ascriptions: Mapping measures of the self ☆

Dan Zahavi [a,*], Andreas Roepstorff [b]

[a] Center for Subjectivity Research, Department of Media, Cognition and Communication, University of Copenhagen, Denmark
[b] Department of Social Anthropology & Center for Functionally Integrative Neuroscience, Aarhus University, Denmark

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Self
Brain imaging
Adjective ascription task
Mirror self-recognition
Experimental psychology
History of psychology (science)
Philosophy
Interdisciplinary research
Face recognition
Self-awareness

ABSTRACT

The 'self' is increasingly used as a variable in cognitive experiments and correlated with activity in particular areas in the brain. At first glance, this seems to transform the self from an ephemeral theoretical entity to something concrete and measurable. However, the transformation is by no means unproblematic. We trace the development of two important experimental paradigms in the study of the self, self-face recognition and the adjective self ascription task. We show how the experimental instrumentalization has gone hand in hand with a simplification of the self-concept, and how more conceptual and theoretical reflections on the structure, function and nature of self have either disappeared altogether or receded into the background. We argue that this development impedes and complicates the interdisciplinary study of self.

1. Introduction: conflicting perspectives on self

The nature, structure and reality of self have been discussed by philosophers for millennia. In the current debate, a form of self-skepticism is not uncommon. As Thomas Metzinger puts it at the very beginning of his 2003 book *Being No One*: "...no such things as selves exist in the world. Nobody ever *was* or *had* a self" (Metzinger, 2003, p. 1). In defending this view, Metzinger, however, endorses a rather reified notion of self. If it exists, the self must be a non-physical soul-substance. Metzinger denies the existence of such an unchangeable and ontologically independent entity, and therefore argues that the self is illusory (Metzinger, 2003, pp. 370, 385, 390). Metzinger's conclusion, however, is only warranted if his definition of self is the only one available, but that is hardly the case. Quite to the contrary, in fact, since most empirical researchers who currently investigate the development, structure, function and pathology of self, usually employ and operate with quite different notions of self. If one thought that only philosophers would be interested in investigating the nature and existence of self and that scientists would stay away from a topic as elusive as this, one is consequently bound to be surprised. If anything, recent years have witnessed a dramatic increase of interest in self in disciplines as various as cognitive science, developmental psychology, sociology, neuropsychology and psychiatry. Consider, for instance, Neisser's already classical distinction between the ecological, the interpersonal, the extended, private and conceptual self (Neisser, 1988). Consider the interest in self found in emotion research. As a recent text book puts it: "One cannot study self-conscious emotions without trying to conceptualize the self and its many levels and its role in the generation of emotions" (Campos, 2007, p. xi). Consider, for a final example among many, how current work on pathologies as diverse as schizophrenia and dementia refer to and discuss

☆ This article is part of a special issue of this journal on Brain and Self: Bridging the Gap Special Issue: T. Feinberg.

* Corresponding author. Address: Center for Subjectivity Research, University of Copenhagen, Njalsgade 140-142, 5th floor, DK-2300 Copenhagen S, Denmark. Fax: +45 35328681.
E-mail address: dza@hum.ku.dk (D. Zahavi).

1053-8100/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.concog.2010.10.011

Please cite this article in press as: Zahavi, D., & Roepstorff, A. Faces and ascriptions: Mapping measures of the self. *Consciousness and Cognition* (2010), doi:10.1016/j.concog.2010.10.011

the topic of self. As Seeley and Miller wrote in 2005: "Though once relegated to philosophers and mystics, the structure of the self may soon become mandatory reading for neurology, psychiatry, and neuroscience trainees. For the dementia specialist the need for this evolution is transparent, as shattered selves – of one form or another – remain a daily part of clinical practice" (Seeley & Miller, 2005, p. 160). One might wonder how Seeley and Miller would react to Metzinger's claim that there never has existed anybody who was or had a self.

But what should one conclude from this discrepancy between the perspectives on self found in contemporary philosophy and empirical research? If we reconsider Seeley and Miller's assertion it can obviously be interpreted in two very different ways. Is the idea that empirical researchers should become familiar with philosophical discussions of self, since the latter are of relevance for the empirical research, or is the idea rather that empirical researchers should take on the task of analyzing and explaining the self themselves? It is certainly not difficult to find vocal representatives of the latter view. Crick, for instance, has argued that it is hopeless to try to solve the problems of self and consciousness by general philosophical arguments. In his view, what we need are suggestions for new experiments that might clarify and ultimately solve these problems (Crick, 1995, p. 19). Indeed, on Crick's view, "the study of consciousness is a scientific problem. [...] There is no justification for the view that only philosophers can deal with it" (Crick, 1995, p. 258). Quite on the contrary, in fact, since philosophers "have had such a poor record over the last two thousand years that they would do better to show a certain modesty rather than the lofty superiority that they usually display" (Crick, 1995, p. 258).

Crick's somewhat polemical assertions do raise a fundamental question. What is the right way to conceive of the relationship between philosophical analysis and empirical investigation when it comes to the study of self? One tempting reply might be that the task left to philosophy is to pose the questions and that empirical science can then provide the answers. This cannot quite be right, however, if only because the alleged empirical answers, as we shall soon see, are often in urgent need of conceptual clarification.

Not surprisingly, one of the main challenges to a neuroscientific investigation of self has been to identify and locate the neural correlate of self. In a survey article entitled "Is self special: a critical review of evidence from experimental psychology and cognitive neuroscience" which was published in *Psychological Bulletin* in 2005, Gillihan and Farah, two neuroscientists, discussed the different suggestions that neuroscience had recently been offering. Their conclusion was somewhat discouraging in that different researchers had pointed to quite different areas in the brain.

What is the reason for this lack of consensus? There are several different ones. But let us focus on a single, which is of particular relevance in this context. When one reads research on self written by neuroscientists much effort is usually spent on explaining the experimental setup and on discussing and interpreting the experimental results. Much less time is devoted to discussing and clarifying the very notion of self at work. For one example among many, consider a 2005 piece by Baron-Cohen, where he writes: "I do not tackle the thorny question of how to define the self [...]. Rather, I accept that this word refers to something we recognize and instead raise the question: are people with autism trapped – for neurological reasons – to be totally self-focused?" (Baron-Cohen, 2005, p. 166). But does it really make sense to discuss whether autism involves a disturbed focus on self, if one does not spend any time discussing and defining the concept of self at play? Perhaps one could object that the notion of self is so univocal and obvious that it is superfluous with a more thematic demarcation and clarification, but that retort is easy to dismiss. The current discussion of self is quite diversified, to put it mildly. Given this situation, it also does not make much sense – apropos the survey article of Gillihan and Farah – to discuss where the neural correlate of self is located if one does not at the same time make it clear, what concept of self one is operating with, as well as make it clear why one takes one's point of departure in precisely this concept rather than in another concept. Indeed, part of the lack of consensus documented by Gillihan and Farah might precisely be due to the fact that various experimentalists operate with different notions of self.

This is per se not a bad thing – indeed on our view, the self is so multifaceted a phenomenon that various complementary accounts must be integrated if we are to do justice to its complexity. But needless to say, this complexity necessitates conceptual clarification, since a lack of clarity in the concepts used will lead to a lack of clarity in the questions posed, and thus also to a lack of clarity in the design of the experiments supposed to provide an answer to the questions. In the following, we will exemplify this lack of conceptual clarity, but we will also show that as different notions of self become instrumentalized in particular experimental designs, there is a risk that the initial precision and care with which the notions of self were introduced and defined becomes discursively buried under particular developments in experimental design. We shall demonstrate that this allows for a slip, over time as research progresses, from the experiment as a model *of* a particular notion of the self to the experiment as a model *for* Self (Geertz, 1966). This shift does not only constitute a category mistake, it also further complicates necessary interdisciplinary work on these issues.

To quickly outline the structure of our contribution: We first intend to look more carefully at two highly influential paradigms in the neuroscientific investigation of self – the study of facial self-recognition and the study of adjectival self-attribution. Tracing the history of these paradigms clarifies how they have developed from highly particular, and highly different, theoretical notions of what the self might be. However, as these approaches have become increasingly instrumentalized in concrete experiments, the premises, which came with the original positions, have increasingly become implicit. Following these trajectories of scientific practices may thus make explicit tacit assumptions and limitations that characterize each paradigm. This will not only show how cautious one ought to be when making claims about neuroscientific solutions to the problem of self, but also exemplify why more theoretical reflections will remain of paramount importance in this ongoing enterprise.

Please cite this article in press as: Zahavi, D., & Roepstorff, A. Faces and ascriptions: Mapping measures of the self. *Consciousness and Cognition* (2010), doi:10.1016/j.concog.2010.10.011

2. The face of the self

Let us begin with the neuroscientific study of facial self-recognition. The assumption has typically been that if there are areas in the brain that show more pronounced activity when one recognized one's own face (compared to what happens when one recognizes other familiar faces) then the respective areas in the brain must constitute or at least be a central part of the neural correlate of self (Gillihan & Farah, 2005). We wish to probe and highlight two main problems and limitations with this approach. The first question concerns whether the recognition of a visual representation of one's own face should count as a paradigmatic and fundamental instance of self-experience. The second question is whether the form of self-experience that facial self-recognition arguably does exemplify is as socially and culturally impoverished as it has often been made out to be.

In articles and books such as "Where in the brain is the self?", "Where am I? The neurological correlates of self and other", and *The Face in the Mirror: How we know who we are* Keenan's (and colleagues') search for the neural location of self has led to the study of facial self-recognition. One of the recurrent findings reported by Keenan is right frontal lateralized activation for self-face recognition – there is more than twice the activity for self-faces compared to familiar faces (Feinberg & Keenan, 2005, p. 673) – and Keenan has claimed that this empirical evidence provides support for the right hemisphere model of self-awareness (Platek, Keenan, Gallup, & Mohamed, 2004, p. 119). When reading the various publications, one is immediately struck by the almost complete absence of an actual working definition of both self and self-awareness. In *The Face in the Mirror*, however, it is acknowledged that a definition might be in place, and it is proposed that self-awareness amounts to higher-order consciousness and metacognition (Keenan, Gallup, & Falk, 2003, pp. xi–xi, 54, 57). At the same, however, it is also stated that consciousness might be used as a synonym for self-awareness (Keenan et al., 2003, pp. xix, xxi). As we will see in a moment, these initial definitions are far from innocuous.

One question to ask though is why self-face recognition is considered particular relevant and important? Why does it tell us something important about self? It is not difficult to see that Keenan's investigation of facial self-recognition is indebted to an older and still highly influential paradigm in developmental psychology and comparative psychology, namely the attempt to subject children, chimpanzees, elephants, dolphins and most, recently, magpies to the mirror self-recognition task in order to test for the presence of self-awareness. Indeed this debt is explicitly admitted by Keenan, who points to Gallup's classical work, in particular his 1970 and 1982 articles, for providing the theoretical framework (Platek et al., 2004, p. 114).

Why did Gallup originally attribute such importance to the passing of the mirror self-recognition task? The standard answer is that the exhibition of self-directed behavior toward a mark surreptitiously put on the face and discovered in the mirror provides empirical and operational evidence for the presence of conceptual self-awareness (Gallup, 1977, p. 337). On closer look, however, it turns out that Gallup was not merely stressing the link between mirror self-recognition and self-awareness. He also took the passing of the mirror task to be a litmus test for the possession of consciousness. Thus, on Gallup's view, consciousness is bidirectional. It allows one to attend outwardly to things in the world, but also to attend inwardly and to monitor one's own mental states (Gallup, 1982, p. 242). To that extent, consciousness covers and includes both awareness and self-awareness. In continuation of this line of thought, Gallup also claimed that conscious experience necessarily presupposes self-awareness and that creatures that lacked the ability to monitor their own mental states were mindless. Either one is aware of being aware, or one is unaware of being aware, and the latter amounts to being unconscious (Gallup, 1982, pp. 243, 245; Gallup, 1985, p. 638). Following this line of reasoning, Gallup concluded that although most organisms behave *as if* they are conscious and minded, prior to the emergence of self-awareness as evidenced from their ability to pass the mirror self-recognition task, they are mindless. They lack conscious experience, and only possess unconscious sensations, pains, etc. (Gallup, 1982, p. 242; Gallup, 1985, p. 638).

Although Keenan does not explicitly endorse such a view, his own definitions seem to point in similar directions. When defining self-awareness in terms of reflective metacognition, and when saying that consciousness is a synonym for such self-awareness, it follows that creatures incapable of such reflective metacognition also lack consciousness. To put it differently, Keenan's initial definition ultimately commits him to a highly controversial higher-order representational account of consciousness according to which a creature only enjoys conscious experiences if it has the ability to reflect upon its own mental states.

There are many good reasons for resisting such a conclusion, which has rather dramatic implications not only for our ascription of conscious experiences, i.e., mental episodes with phenomenal feels, to infants, but also to all those animals which lack a cognitive capacity for reflection. In fact, higher-order representational accounts of consciousness have come under increasing attack in recent years (cf. Kriegel & Williford, 2006; Zahavi, 2004, 2005), and perhaps Keenan himself would refrain from such a view. At least, they are not implications he explicitly draw and endorse. So let us return to his central claim. Keenan repeatedly claims that the ability to pass the mirror self-recognition test demonstrates the capacity for self-awareness. But how strong is this correlation supposed to be? There is hardly any doubt that Keenan opts for a strong claim, which would also match Gallup's view on the matter. This is why Keenan claims that the passing of the mirror test is highly correlated with every indicator of self-awareness and that the absence of mirror self-recognition is correlated with an absence of other self-aware behaviors (Keenan et al., 2003, p. 22). It is not entirely clear, however, that Keenan's own findings really support these statements. Take the case of people suffering from the delusional misidentification symptom known as mirror sign, which Keenan also discusses. Presumably they are incapable of passing the mirror self-recognition test – and the reason they have this incapacity is specifically related to problems with self-recognition, and is not simply due to some form

Please cite this article in press as: Zahavi, D., & Roepstorff, A. Faces and ascriptions: Mapping measures of the self. *Consciousness and Cognition* (2010), doi:10.1016/j.concog.2010.10.011

of prosopagnosia. But if one were to take Keenan seriously, people such as these, would also lack self-awareness and the capacity for self-reflection. But there is no evidence that this happens to be the case.

More generally speaking, it is not difficult to come up with quite general objections to the idea that our ability to identify a visual representation of our own face should constitute a particular central or fundamental form of self-awareness. Although facial self-recognition might testify to the existence of a form of self-awareness, the failure to recognize one's own face certainly does not prove the absence of every form of self-awareness. To put it differently, the absence of facial self-recognition might be perfectly compatible with the presence of other forms of self-awareness. Indeed, as we see it, one decisive problem facing Keenan's interpretation of his own findings is that he underestimates how complex and varied self-experience is. Not only does he disregard the possibility that phenomenal consciousness involves self-awareness in the weak sense that there is something it is like for the subject to have the experience, i.e., that the distinct first-personal character of phenomenal consciousness amounts to a low-level form of self-awareness (cf. Flanagan, 1992; Zahavi, 1999, 2003, 2005), but he also seems to ignore the possibility that infants might have a sense of their own bodies as organized and environmentally embedded entities long before they are able to pass any mirror self-recognition tasks, and, hence, an early embodied sense of themselves in perception and action. Thus, as many developmental psychologists have pointed out, already from around 3 months of age, infants discriminate what pertains to the self and what pertains to someone else interacting with them. In the footsteps of Neisser and Gibson, one could call this early sense of self the infant's ecological self (Rochat, 2001, pp. 30–31, 41). Would we really be able to recognize our own mirror-image, which presumably relies on a detection of the perfect cross-modal match between our own bodily movements and the movements of the mirror-image, if we were not already proprioceptively aware of our *own* bodily movements? To put it differently, would one really be able to recognize oneself in the mirror, if one lacked bodily self-awareness? There are even those who claim that this bodily self-awareness constitutes the fundamental requirement. This is for instance the claim made by Mitchell in a number of papers. He has argued that mirror self-recognition merely requires a kinesthetic sense of own body (subjective self-awareness), a capacity for kinesthetic-visual matching and an understanding of mirror-correspondence (Mitchell, 1997a, p. 31; Mitchell, 1997b, p. 41). If so, it would obviously invalidate Gallup and Keenan's claim that mirror self-recognition requires both introspection and the possession of a self-concept or an internal self-model (cf. Keenan et al., 2003, p. 11). Indeed, as Mitchell has pointed out, it remains quite unclear what mental state a creature is supposed to attend to in recognizing itself in the mirror (Mitchell, 1997a, p. 23).

All of this is not to say that mirror self-recognition is insignificant, the question though is whether Keenan and before him Gallup have realized its proper significance. In some of his early writings, Gallup took mirror self-recognition to testify to the perfect match between the observer and the observed. As he put it "The unique feature of mirror-image stimulation is that the identity of the observer and his reflection in a mirror are necessarily one and the same" (Gallup, 1977, p. 334). In addition, he repeatedly emphasized the distinction between social responsiveness and self-directed mirror behavior and claimed that one in recognizing one's mirror-image ceased to respond socially to it (Gallup, 1970, p. 86). But is this really correct? Consider that a visual representation of one's own face provides one with information about oneself that is very different from what one hitherto has been in possession of. Consider that the face we see reflected in the mirror is also the face others see when we interact with them. Indeed one of the reasons why people spend so much time before a mirror engaged in impression management is precisely because of the high social valence of one's face. When seeing myself in the mirror, I am confronted with the appearance I present to others. To see oneself in the mirror (or in a photo) is to become a spectator of oneself. It is to adopt a perspective or viewpoint on oneself that equals what others can adopt. Contrary to what Gallup is claiming, to recognize oneself in the mirror does not simply involve an identification of the felt me which is here, and the perceived me which is there. There is more at stake than a simple affirmation of a pre-existing identity attached to it, there is also what Merleau-Ponty described as the unsettling experience of realizing that the felt me has an exterior dimension that can be witnessed by others (cf. Merleau-Ponty, 1964, pp. 129, 136, 140). There is an awareness of one's public appearance. In fact, not only am I seeing myself as others sees me, I am also seeing myself as if I was an other, i.e., I am adopting an alienating perspective on myself. To put it differently, the mirror and the photo afford quite new possibilities of adopting an objectified stance towards ourselves. To recognize an image of oneself is to appropriate an objectification of oneself.

More needs to be said in defending this claim. In particular, one has to consider the relation between successful mirror self-recognition in humans and in non-human animals. It is, however, not at all obvious that the two can be equated and that mirror self-recognition in chimpanzees or magpies corresponds to the cognitive and affective self-consciousness manifested in children passing the test (cf. Rochat & Zahavi, in press). For now, however, all we wish to deny is that the recognition of a visual representation of one's own face counts as the paradigmatic and fundamental instance of self-experience and that it is as socially and culturally impoverished as it has often been made out to be. To recognize one's own mirror-image is definitely not to cease responding socially to it. To put it differently, to test facial self-recognition is not to test the self per se, but to test and probe a quite specific dimension of self, e.g., the self as social object. This angle and limitation is something that has not been sufficiently considered in recent neuroscience.

One possible objection to this criticism might run as follows: In both Gallup and Keenan one finds occasional positive references to the work of Mead and Cooley (cf. Gallup, 1977, p. 335; Keenan et al., 2003, p. 41), who both explicitly and persistently discussed the self as social object. It is consequently important to notice that whereas Keenan's work is indebted to Gallup's, Gallup's own account of self-recognition is precisely influenced by ideas found much earlier in Cooley and Mead (cf. Gallup, 1975, 1983; Gallup, McClure, Hill, & Bundy, 1971). However, on closer examination it turns out that Gallup in the process of developing his own account gradually altered and ultimately inversed these ideas (for a careful analysis of Gallup's

Please cite this article in press as: Zahavi, D., & Roepstorff, A. Faces and ascriptions: Mapping measures of the self. *Consciousness and Cognition* (2010), doi:10.1016/j.concog.2010.10.011

puzzling reliance and use of these authors, see Mitchell, 1997a). Whereas Cooley and Mead argued that self-knowledge presupposes knowledge of others and that the self-concept derives from one's taking the perspective of another toward oneself (Cooley, 1912, 246; Mead, 1962, 138), Gallup and Keenan both defend the view that knowledge of others presupposes knowledge of self and a developed self-concept. Gallup writes that the tendency to impute mental states to others presupposes the capacity to monitor such states on the part of the individual making the imputation (Gallup, 1982, 243), and Keenan argues that it is because I know my own thoughts that I can predict or infer another person's mental state (Keenan et al., 2003, p. 78).

As should have become clear by now, the neuroscientific investigation of facial self-recognition relies on a theoretical framework with a long (and partly forgotten) history. When it comes to its more principled reflections on and analyses of central concepts such as consciousness, self-awareness and self – which shaped the design of the experiment and continue to influence the interpretation of the empirical findings – these are clearly inadequate and in many ways committed to rather controversial theses.

3. Ascribing myself

In recent brain imaging literature, arguably the most influential experimental paradigm for identifying putative neural correlates of the self has been the adjective ascription task. Briefly, the task involves placing experimental subjects in a suitable brain scanner and presenting them with a list of adjectives. While their brains are being scanned, subjects have to evaluate the list of adjectives in different ways, one of which involves evaluating whether the word appropriately describes themselves. In this way, it is possible to keep the visual input constant between the different sessions, while the experimental contrast is provided by the script, which specifies what the experimental subject should do with the word (Jack & Roepstorff, 2002). The task was used both in an early PET study by Craik and colleagues (1999) and in an early fMRI study by Kelley and colleagues (2002). Both papers have been highly influential (with 346 citations and 535 citations respectively according to Google scholar at the time of writing).

While the title of the Craik et al., study *In Search of the Self: A Positron emission tomography study* suggested an open ended exploration, by the time of the Kelley paper *Finding the Self? An event related fMRI study* only the question mark, highly unusual for the title of a neuroscience paper, suggested some uncertainty. Effectively, the paradigm today provides a ready-made technology to identify a neural correlate of the self; you take a scanner, add an adjective ascription task, and compare self ascription to e.g. other ascription or mother ascription (Ray et al., 2010; Vanderwal, Hunyadi, Grupe, Connors, & Schultz, 2008). Then one does the trick that brain scanners can do (Roepstorff, 2002), and out comes a set of data that may be transformed to an image (Roepstorff, 2007). Just as with the mirror-recognition task, one may ask the question why is adjective ascription considered a particularly relevant task to identify neural correlates of the self? To the extent that the self, at least analytically, effectively becomes what the experiment measures, it may be instructive to follow in some detail the construction of this paradigm for eliciting the self.

The origin of the self ascription task is to be found more than 30 years ago, in a memory study that investigated whether self-reference could serve a function in the processing of certain kinds of information (Rogers, Kuiper, & Kirker, 1977). Their experiment combined two lines of experimentation. One of them was the already then classic idea that schemata or prototypes are key in organizing memory (Bartlett, 1932). A number of researchers in social and personality psychology had recently suggested that personality traits in general could be conceived of as such schemata. This had been tested e.g. in how subjects would remember trait adjectives about characters in a narrative (see e.g. Cantor & Mischel, 1977).

Rogers et al. extrapolated this idea to the self, and suggested that self-reference could involve a schema of trait-like features, abstracted over the life history of the individual, which could be seen as a feature list. Hence, "[w]hen self-reference is involved, it should provide a useful device for encoding or interpreting incoming information by virtue of accessing the extensive past experience abstracted in the self" (Rogers et al., 1977, 678–9). They then combined this idea of the self with recent experimental work on memory retention, which found that words, which a subject had been exposed to, would be remembered differently, depending on how *deep* the processing had been. Craik and Tulving (1975) had recently demonstrated that words presented as part of a *semantic task*, which probed their meaning, were remembered better than words presented as part of a *phonemic task*, which probed the letters in the words, and these words were, again, recalled better than words presented in a *structural task* where subjects had to decide if the words had been written in large or in small letters. These data had been interpreted as support for the position that the strength of the memory trace is "a positive function of 'depth' of processing, where depth refers to greater degrees of semantic involvement" (Craik & Tulving, 1975 p. 268, cited from Rogers et al., 1977).

With these two lines of experimentation in hand, Rogers et al. had an easy way to test the hypothesis that if the self was a prominent prototype in organizing experiences, then relating words to 'the self' would be a very efficient way to remember them. In their own words: "If the self is an active agent in the encoding of personal data, we predicted that the self-reference rating would produce good incidental recall in this depth-of-processing paradigm. If incidental recall of the self-reference words is superior to that for semantic words, the hypothesis that the self serves an active and powerful role in processing personal data would be supported" (Rogers et al., 1977, p. 680).

The design of the study was quite straight forward. It took the design developed by Craik and Tulving (1975) and added a 'self' related component. That is, subjects saw lists of adjectives on four tasks designed to force varying kinds of encoding: structural, phonemic, semantic, and self-reference, and they should consider whether each word fittingly described them.

Please cite this article in press as: Zahavi, D., & Roepstorff, A. Faces and ascriptions: Mapping measures of the self. *Consciousness and Cognition* (2010), doi:10.1016/j.concog.2010.10.011

After this encoding phase, subjects incidental recall of the rated words were examined. This indicated, in accordance with the stated hypothesis, that adjectives rated under the self-reference task were recalled better than semantically, phonetically and structurally encoded words. The authors hence concluded that "the present data indicate unequivocally that words rated under the self-reference task show superior recall. This indicates that self-reference represents a powerful and rich encoding device" (Rogers et al., 1977, p. 685). Further, "In order for self-reference to be such a useful encoding process, the self must be a uniform, well-structured concept. During the recall phase of the study, subjects probably use the self as a retrieval cue [...]. In order for this to be functional, the self must be a consistent and uniform schema [...]. In sum, the self contains a set of ordered features [...]. The ordering appears to be from general to specific, with the general terms (e.g., traits) ordered by a combination of salience and extremity. The general terms can serve as schemata when studied independently of a person's idiographic view of self" (op. cit., p. 686).

The article is a masterpiece in experimental psychology, and it demonstrates quite clearly a particular *style* of producing facts (Fleck, 1979). Through a combination of different pieces of experimental work, that each solves a particular problem, bit-by-bit a new argument is constructed, almost like constructing out of Lego pieces a particular building. In the same manner, this adjective ascription task, first used to test a particular version of how relating to the self may allow a particular 'deep' level of processing, has now become used as a proxy to identify in the brain "the neuronal correlate of the self" (Craik et al., 1999; Kelley et al., 2002). These and other studies identified a number of regions, but in particular there seemed to be more activity in the so-called medial prefrontal cortex (MPFC) during the 'self' part of the paradigm across a number of adjective ascription experiments. Once such a putative 'self' area was found, differential activity could be, and has been, used to study putative cultural (Zhu, Zhang, Fan, & Han, 2007) or religious (Han et al., 2008) aspects of the self, in relation to differences in self-construal within a given culture (Ray et al., 2010), between different generations (e.g. Feyers, Collette, D'Argembeau, & Majerus, 2010; Gutchess, Kensinger, & Schacter, 2010), between individuals with autism and a control group (Kennedy & Courchesne, 2008) or persons with depression and a control group (Lemogne et al., 2009) or as dynamic modulations of culturally mediated perceptions of the self (Chiao et al., 2010).

A striking development happened in this transformation from a memory task to a technology to identify different aspects of the self between different groups. This occurred over the course of more than 30 years and through a set of subsequent experimental steps that all involve amplifications and reductions of specific features (Latour, 1999; Roepstorff, 2004). The first experiments began with the assumption that personal traits could be seen as prototypes. This led to a 'self as traits' approach, where the self appeared to allow for deep processing. Once taken to the scanner environment, the data were probed to see if any parts of the brain could be indicative of a 'deep' processing, indexed by more activity in these regions. The next step was the assumption that as these regions appeared to be more involved in self processing, activity in these regions could provide a 'neural signature' of the self and they could be seen as 'self related' areas. If that were the case, activity in these regions could be interpreted as an index of self-related activities. Therefore, differences in activity during self related tasks in these areas could be a sign of differences in self processing between groups. Hence, differences in 'self' could be a trait for that group, and that trait could, perhaps, be considered prototypical. This progression has been a very powerful fact making strategy, but it also appears to have an inherent circularity that may be unfolded in the following way:

> personal trait as prototypes → self as traits → self as deep processing → deep-processing as neuronally located → neural location as index of self → neuronal level of activation in groups as index of self → self as group trait → group trait as prototype for the individual self

In other words, what began as a particular conceptualization of the self as a schema, which was handy because it could easily be implemented in an experiment, ended with subsequent generations of experiments purporting to provide a model for where the self is in the brain, and ultimately what the self is. This shift from a model *of* the self to a model *for* the self demonstrates, we believe, that the paradigms remain dependent upon the underlying assumptions, i.e., they point to and rely on a particular understanding of the self. This is very clearly seen in the early papers, where the paradigms were developed. The authors were acutely aware that the argument was, in the end, conceptual. However, as the paradigms moved across disciplines and techniques, this understanding became buried under new shifts in perspectives, new problems to solve. Whereas in the beginning the aim had been to design experiments that could allow one to get a handle on what a self is, in the end the self ended up being defined in terms of what the experiment could handle.

As also pointed out by Gillihan and Farah (2005) in their very careful review, the observation of apparent differential activity in the medial prefrontal cortex during a self ascription task cannot be considered a 'proof' that the 'self' is located in these areas. Remember the origin of the task, it began as a memory experiment that investigated whether adjectives were remembered better if they were related to the self. This was based on the argument that 'the self' formed a strong schema for categorizing and remembering incoming information. Given that the knowledge of oneself is probably better than the knowledge of abstract others such as the former US president (Kelley et al., 2002) or the Danish Queen (Lou et al., 2004) it cannot be ruled out that the differential activity in MPFC is a memory effect which reflects a difference in familiarity. This may explain, as Gillihan and Farah (2005) also suggests, that in some experimental contexts, close others and the self both appear to activate this region (see e.g. Zhu et al., 2007 & Vanderwal et al., 2008 for 'mother related' case). In this alternative interpretation, which closely follows the original rationale behind the self ascription task, close others, like oneself, may be used as powerful schema for organizing and remembering personal adjectives. It is obviously interesting if between individuals or groups, no matter whether these are 'clinical' categories or cultural' categories, different types of persons form such

Please cite this article in press as: Zahavi, D., & Roepstorff, A. Faces and ascriptions: Mapping measures of the self. *Consciousness and Cognition* (2010), doi:10.1016/j.concog.2010.10.011

schema. This suggests different patterns in how self-other relations are constituted, and examining how such relations are changed, e.g. by priming, may point to important contextual dynamics both of the self and of the brain (Ng, Han, Mao, & Lai, 2010). Such a line of research provides a potentially powerful challenge to universalizing tendencies in the cognitive sciences (Roepstorff, in press), and it obviously differs considerable from the claim that the self is localized somewhere in the brain and that activity in this region is a sign of a particular form of selfhood.

4. Conclusion

In various publications Keenan and colleagues have claimed that the search for the localization of the self in the brain has been the goal of consciousness research for centuries (Feinberg & Keenan, 2005, p. 661), and that this problem remains one of the great mysteries of science, philosophy and psychology (Keenan et al., 2003, p. 99). In the article entitled "Where in the brain is the self?" which Keenan co-authored with Feinberg, the authors are careful in modifying the initial claim by conceding not only that modules of the brain do not exist in isolation, and that one has to view the brain in its entirety (Feinberg & Keenan, 2005, p. 673), but also that it might be more appropriate and correct to opt for the more modest claim that the right hemisphere is dominant for certain aspects of self, than to make the stronger claim that the self resides in the right hemisphere (Feinberg & Keenan, 2005, p. 675). We could not agree more. It is indeed far better to label the search for the neural correlates of self, a search for those neural structures and mechanisms that enable self-recognition and self-experience, than to describe it as an attempt to locate the self in the brain. The latter claim is in our view (and here we would side with so otherwise antagonistic philosophers as Dennett and Hacker) tantamount to a category mistake (Bennett & Hacker, 2003; Dennett, 1992). For that very reason, the answer to the question, where in the brain is the self, can only be nowhere. To say that the self is nowhere in the brain, is, however, not to say that "nobody ever was or had a self" (Metzinger, 2003, p. 1). We are not self-skeptics. We do not deny the reality or question the ontological status of the self (cf. Siderits, Thompson, & Zahavi, 2010, 2005), we are simply denying that the self is a kind of thing that can be found in the brain.

Again, this is by no means to question the value of some of the existing paradigms. In fact, a task like the adjective ascription task is a simple, effective and easily implementable method that has in the past taught us much about how memories are stored and processed, and used in combination with brain scanners may tell a lot about how probing knowledge, affects and expectations about oneself and about others may tie into particular networks of the brain. Equally, differential activities in these networks across groups, contexts and primings may give important hints both to the workings of the brain and of the mind. However, we are not convinced that this translates into the self being 'located' in those parts of the brain that are activated by a self-directed adjective ascription task. Moreover, as our analysis of facial self-recognition ought to have shown, neuroscientific research on the neural correlates self-recognition and self-experience necessitates rather than obviates the need for a careful conceptual analysis. Even recognizing something as apparently simple as an image of oneself is more complicated than just contrasting self and other. It involves the appropriation of an objectification of oneself, and thus entails a critical tension between experiencing oneself as object and experiencing oneself as subject (Legrand, 2007). This is nothing but a special case of a much more complicated pattern of how people in interaction with others constitute and develop themselves, and others (Zahavi, 2009, 2010). Such intersubjective interplay can be followed even in concrete interactions during a simple cognitive experiment (Roepstorff, 2001).

The self is complex and multidimensional. This complexity necessitates interdisciplinary collaboration; collaboration across the divide between theoretical analysis and empirical investigation. To think that a single discipline, be it philosophy or neuroscience, should have a monopoly on the investigation of self is merely an expression of both arrogance and ignorance.

Acknowledgments

The authors acknowledge the financial support from the Danish National Research Foundation and from the Danish Council for Independent Research: Humanities.

References

Baron-Cohen, S. (2005). Autism-'Autos': Literally, a total focus on the self? In T. E. Feinberg & J. P. Keenan (Eds.), *The lost self: Pathologies of brain and identity* (pp. 166–180). Oxford: Oxford University Press.
Bartlett, F. C. (1932). *Remembering. A study in experimental and social psychology*. Cambridge: Cambridge University Press.
Bennett, M. R., & Hacker, P. M. S. (2003). *Philosophical foundations of neuroscience*. Oxford: Blackwell.
Campos, J.J. (2007). Foreword. In J. L. Tracy, R. W. Robins, & J. P. Tangney (Eds.), *The self-conscious emotions: Theory and research* (pp. ix–xii.). New York: The Guilford Press.
Cantor, N., & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. *Journal of Personal and Social Psychology, 35*, 38–48.
Chiao, J., Harada, T., Komeda, H., Li, Z., Mano, Y., Saito, D., et al (2010). Dynamic cultural influences on neural representations of the self. *Journal of Cognitive Neuroscience, 22*, 1–11.
Cooley, C. H. (1912). *Human nature and the social order*. New Brunswick, NJ: Transaction Books.
Craik, F., Moroz, T., Moscovitch, M., Stuss, D., Winocur, G., Tulving, E., et al (1999). In search of the self: A positron emission tomography study. *Psychological Science, 10*, 26–34.
Craik, F., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268–294.
Crick, F. (1995). *The astonishing hypothesis*. London: Touchstone.

Please cite this article in press as: Zahavi, D., & Roepstorff, A. Faces and ascriptions: Mapping measures of the self. *Consciousness and Cognition* (2010), doi:10.1016/j.concog.2010.10.011

Dennett, D. C. (1992). The self as the center of narrative gravity. In Frank S. Kessel, Pamela M. Cole, & Dale L. Johnson (Eds.), *Self and consciousness: Multiple perspectives* (pp. 103–115). Hillsdale, NJ: Erlbaum.
Feinberg, T. E., & Keenan, J. P. (2005). Where in the brain is the self? *Consciousness and Cognition, 14*, 661–678.
Feyers, D., Collette, F., D'Argembeau, A., & Majerus, S. (2010). Neural networks involved in self-judgement in young and elderly adults. *Neuroimage, 53*, 341–347.
Flanagan, O. (1992). *Consciousness reconsidered*. Cambridge, MA: MIT Press.
Fleck, L. (1979). *Genesis and development of a scientific fact*. Chicago: Chicago University Press.
Gallup, G. G. (1970). Chimpanzees: Self-recognition. *Science, 167*, 86–87.
Gallup, G. G. (1975). Towards an operational definition of self-awareness. In R. H. Tuttle (Ed.), *Socioecology and psychology of primates* (pp. 309–341). Paris: Mouton & Co.
Gallup, G. G. (1977). Self-recognition in primates: A comparative approach to the bidirectional properties of consciousness. *American Psychologist, 32*, 329–338.
Gallup, G. G. (1982). Self-awareness and the emergence of mind in primates. *American Journal of Primatology, 2*(3), 237–248.
Gallup, G. G. (1983). Toward a comparative psychology of mind. In R. L. Mellgren (Ed.), *Animal cognition and behavior* (pp. 473–510). Amsterdam: North-Holland Publisher.
Gallup, G. G. (1985). Do minds exist in species other than our own? *Neuroscience and Biobehavioral Reviews, 9*(4), 631–641.
Gallup, G. G., McClure, M. K., Hill, S. D., & Bundy, R. A. (1971). Capacity for self-recognition in differentially reared chimpanzees. *Psychological Record, 21*, 69–74.
Geertz, C. (1966). *Religion as a cultural system. In the interpretation of cultures* (pp. 87–125). New York: Basic Books.
Gillihan, S. J., & Farah, M. J. (2005). Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. *Psychological Bulletin, 131*(1), 76–97.
Gutchess, A., Kensinger, E., & Schacter, D. (2010). Functional neuroimaging of self-referential encoding with age. *Neuropsychologia, 48*, 211–219.
Han, S., Mao, L., Gu, X., Xhu, Y., Ge, J., & Ma, Y. (2008). Neural consequences of religious belief on self-referential processing. *Social Neuroscience, 3*, 1–15.
Jack, A. I., & Roepstorff, A. (2002). Introspection and cognitive brain mapping: From stimulus-response to script-report. *Trends in Cognitive Sciences, 6*, 333–339.
Keenan, J. P., Gallup, G. G., & Falk, D. (2003). *The face in the mirror: How we know who we are*. New York: HarperCollins.
Kelley, W., Macrae, C., Wyland, C., Caglar, S., Inati, S., & Heatherton, T. (2002). Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience, 14*, 785–794.
Kriegel, U., & Williford, K. (Eds.). (2006). *Consciousness and self reference*. Cambridge, Mass.: MIT Press.
Kennedy, D., & Courchesne, E. (2008). Functional abnormalities of the default network during self-and other-reflection in autism. *Social Cognitive and Affective Neuroscience, 3*, 177.
Latour, B. (1999). *Pandora's hope: Essays on the reality of science studies*. Cambridge, Mass.: Harvard University Press.
Legrand, D. (2007). Pre-reflective self-as-subject from experiential and empirical perspectives. *Consciousness and Cognition, 16*(3), 583–599.
Lemogne, C., Le Bastard, G., Mayberg, H., Volle, E., Bergouignan, L., Lehéricy, S., et al (2009). In search of the depressive self: Extended medial prefrontal network during self-referential processing in major depression. *Social Cognitive and Affective Neuroscience, 4*, 305–312.
Lou, H., Luber, B., Crupain, M., Keenan, J., Nowak, M., Kjaer, T., et al (2004). Parietal cortex and representation of the mental self. *Proceedings of the National Academy of Sciences, 101*(17), 6827.
Mead, G. H. (1962). *Mind, self and society. from the standpoint of a social behaviorist*. Chicago: University of Chicago Press.
Merleau-Ponty, M. (1964). The child's relations with others. In J. Edie (Ed.), *The primacy of perception* (pp. 96–155). Evanston, Ill: Northwestern University Press.
Metzinger, T. (2003). *Being no one*. Cambridge, MA: MIT Press.
Mitchell, R. W. (1997a). Kinesthetic-visual matching and the self-concept as explanations of mirror-self-recognition. *Journal for the Theory of Social Behavior, 27*(1), 17–39.
Mitchell, R. W. (1997b). A comparison of the self-awareness and kinesthetic-visual matching theories of self-recognition: Autistic children and others. *Annals of the New York Academy of Sciences, 818*, 39–62.
Neisser, U. (1988). Five kinds of self-knowledge. *Philosophical Psychology, 1*(1), 35–59.
Ng, S., Han, S., Mao, L., & Lai, J. (2010). Dynamic bicultural brains: fMRI Study of their flexible neural representation of self and significant others in response to culture primes. *Asian Journal of Social Psychology, 13*, 83–91.
Platek, S. M., Keenan, J. P., Gallup, G. G., & Mohamed, F. B. (2004). Where am I? The neurological correlates of self and other. *Cognitive Brain Research, 19*, 114–122.
Ray, R., Shelton, A., Hollon, N., Matsumoto, D., Frankel, C., Gross, J., et al (2010). Interdependent self-construal and neural representations of self and mother. *Social Cognitive and Affective Neuroscience, 5*(2–3), 318–323.
Rochat, P. (2001). *The infant's world*. Cambridge, MA: Harvard University Press.
Rochat, Ph. & Zahavi, D. (in press). The uncanny mirror: A re-framing of mirror self-experience. *Consciousness and Cognition*, doi:10.1016/j.concog.2010.06.007.
Roepstorff, A. (2001). Brains in Scanners: An umwelt of cognitive neuroscience. *Semiotica, 134*, 747–765.
Roepstorff, A. (2002). Transforming subjects into objectivity. An ethnography of knowledge in a brain imaging laboratory. *FOLK, Journal of the Danish Ethnographic Society, 44*, 145–170.
Roepstorff, A. (2004). Mapping brain mappers: An ethnographic coda. In R. Frackowiak et al. (Eds.), *Human brain function, 2nd ed* (pp. 1105–1117). London: Elsevier.
Roepstorff, A. (2007). Navigating the brainscape: When knowing becomes Seeing. In C. Grasseni (Ed.), *Skilled Vision. Between apprenticeship and standards* (pp. 191–206). Oxford: Berghahn Press.
Roepstorff, A. (in press). Culture: A site of relativist energy in the cognitive sciences. *Common Knowledge.*
Rogers, T., Kuiper, N., & Kirker, W. (1977). Self-reference and the encoding of personal information. *Journal of Personal and Social Psychology, 35*, 677–688.
Seeley, W. W., & Miller, B. L. (2005). Disorders of the self in dementia. In T. E. Feinberg & J. P. Keenan (Eds.), *The lost self: Pathologies of brain and identity* (pp. 147–165). Oxford: Oxford University Press.
Siderits, M., Thompson, E., & Zahavi, D. (Eds.). (2010). *Self, no self? Perspectives from analytical, phenomenological and Indian traditions*. Oxford: Oxford University Press.
Vanderwal, T., Hunyadi, E., Grupe, D., Connors, C., & Schultz, R. (2008). Self, mother and abstract other: An fMRI study of reflective social processing. *Neuroimage, 41*, 1437–1446.
Zahavi, D. (1999). *Self-awareness and alterity: A phenomenological investigation*. Evanston: Northwestern University Press.
Zahavi, D. (2003). Phenomenology of self. In T. Kircher & A. David (Eds.), *The self in neuroscience and psychiatry* (pp. 56–75). Cambridge University Press.
Zahavi, D. (2004). Back to brentano? *Journal of Consciousness Studies, 11*(10–11), 66–87.
Zahavi, D. (2005). *Subjectivity and selfhood: Investigating the first-person perspective*. Cambridge, MA: The MIT Press.
Zahavi, D. (2009). Is the self a social construct? *Inquiry, 52*(6), 551–573.
Zahavi, D. (2010). Shame and the exposed self. In J. Webber (Ed.), *Reading Sartre: On phenomenology and existentialism* (pp. 211–226). London: Routledge.
Zhu, Y., Zhang, L., Fan, J., & Han, S. (2007). Neural basis of cultural influence on self-representation. *Neuroimage, 34*, 1310–1316.

Please cite this article in press as: Zahavi, D., & Roepstorff, A. Faces and ascriptions: Mapping measures of the self. *Consciousness and Cognition* (2010), doi:10.1016/j.concog.2010.10.011