

Towards an Integrated Theory of Self-Consciousness

Final version published in: *Grazer Philosophische Studien* 84 (2012), pp. 1-12

Self-consciousness is an issue that is of fundamental theoretical significance in philosophy. It is at the root of many other philosophical issues ranging from epistemological questions (i.e., the problem of self-knowledge) to those that are metaphysical (i.e., the concept of a person or a "self") or moral (i.e., issues of moral agency or autonomy). In the philosophy of mind the problem of self-consciousness is closely related to various other philosophical issues, such as cognition, phenomenal consciousness, personal identity and so on.

So far, many attempts have been made to provide a theory of self-consciousness. Following the linguistic turn in analytical philosophy, such a theory is often thought to be found through an analysis of the linguistic expressions of self-consciousness, in particular the semantics of the first person pronoun (Spitzley 2000). However, this approach has been called into question, both by proponents of phenomenological accounts and, more recently, by those advocating theories of nonconceptual content. Proponents of these alternative approaches argue that in order to give an account of self-consciousness we ought to look at nonreflective or nonconceptual ways of representing the self that provide the foundation for our later ability to acquire a concept of ourselves (e.g., Zahavi 2005; Frank 1991; Bermúdez 1998; Hurley 1997; Legrand 2006, 2007; Vosgerau 2009). Accordingly, following this view, self-consciousness can be conceived of as a complex phenomenon that manifests itself in multiple ways. Needless to say, this broader way of addressing the issue raises a number of new questions. For example, how are we to understand the thesis that there are nonconceptual or nonreflective forms of self-consciousness? What is its structure and how does it relate to higher-level forms of self-consciousness?

More generally, the debate on self-consciousness has expanded in scope in recent years and, at the same time, has become more specialised. In addition to the investigation of so-called nonconceptual forms of self-consciousness and their relation to higher-level, conceptual forms of self-representation, other research has emphasized the intersubjective or temporal aspects of self-awareness. Unfortunately, despite intensive research in many of these areas discussions of these different facets of self-consciousness have remained relatively isolated. This may be due in part to the different methodological angles from which they tend to be approached.

The aim of the present special issue of Grazer Philosophische Studien is to address this thematic isolation and methodological fragmentation by bringing together a number of carefully selected articles that explore and bridge the discussion of different facets of self-consciousness. The articles are intended not only to represent the state of the art in their respective areas of research, but also to highlight relations between different research questions as well as different research methodologies, ranging from philosophy of language and mind to phenomenology and cognitive science. In choosing this integrative approach, we hope to bring to the forefront a number of previously underexposed, yet important connections between different approaches.

1. Three facets of self-consciousness

As indicated above, one can identify three general focal points of the current debate: (1) the relationship between conceptual and nonconceptual forms of self-representation; (2) the role of intersubjectivity; and (3) the temporal structure of self-consciousness.

(1) The relationship between conceptual and nonconceptual forms of self-representation

Proponents of theories of nonconceptual content hold that there are ways of representing the world that are independent of the possession of the concepts needed to specify the content of these representations. This view has been applied to the problem of self-consciousness by authors who argue that there are nonconceptual forms of self-representation; candidates for these are, for example, ecological perception, somatic proprioception, or the sense of agency (e.g., Hurley 1997; Bermúdez 1998; Vosgerau 2009). Proponents of this view often make use not only of insights from philosophy of mind and philosophy of language, but also from the empirical sciences, such as developmental psychology or cognitive neuroscience and hence offer interesting connections to these areas of research. In addition, the debate on nonconceptual self-consciousness offers intriguing, albeit often neglected, connections to the notion of pre-reflexive self-consciousness that was developed based on discussions of the work of Fichte (see, e.g., Frank 1991), as well as based on phenomenological investigations of the structure of self-consciousness that have recently gained renewed attention (e.g., Zahavi 2005; Legrand 2006; Crone 2009).

(2) The role of intersubjectivity

An exploration of the ontogenetic roots of our ability for self-conscious thought quickly brings to the forefront the importance of intersubjectivity for the acquisition of this ability. There are different perspectives from which this issue can be discussed. First, phenomenological analyses show that self-consciousness and intersubjectivity are inextricably linked to each other (e.g., Husserl 1973; Zahavi 2001), and that social interaction - in various forms - is an important component in the development of awareness of oneself and of others (e.g. Gallagher & Hutto 2008). Second, insights from cognitive science and developmental psychology suggest that the development of self-consciousness is correlated with the development of a theory of mind (e.g., Happé 2003; Wellman et al. 2001). Third, following Strawson (1959) and Evans (1982), one can argue that the ability to ascribe states and properties to oneself implies the ability to ascribe these also to others who are not oneself.

The relation between intersubjectivity and self-consciousness also offers interesting connections to questions of embodiment and narrative identity.

(3) The temporal structure of self-consciousness

Debates on self-consciousness rarely address its temporal structure.

Self-consciousness, taken as a mental process, is not only a temporally extended episode itself; it also represents the subject's own diachronic persistence (e.g., Bieri 1986; Dainton 2000). Therefore, it would seem that a proper analysis of self-consciousness must take into account its temporal dimension. One way of doing this is by applying the distinction between conceptual and nonconceptual forms of self-consciousness to the representation of the self over time. In a nonconceptual form diachronic self-consciousness can be described as the temporally extended perspective of the experiencing subject, which may be elucidated with the help of phenomenological theories of "inner" time consciousness (Husserl 1985; Zahavi 2003; 2007; Gallagher 2005). In a conceptual form diachronic self-consciousness is narrative in structure and integrates different life episodes allowing for a (more or less) coherent self-understanding (e.g., Schechtman 1996, Gallagher 2007; Hutto 2007). The precise relation between those different modes remains until now rather unexplored.

Focussing on the temporal aspects of self-consciousness allows for promising systematic insights into different recent debates, such as analytical and phenomenological accounts of self-consciousness and time consciousness as well as theories of the transtemporal identity of sentient beings. Another interesting aspect of this debate is the potential role of intersubjectivity for diachronic self-consciousness.

2. Overview of articles

Taking on the epistemological issue of self-knowledge, Dretske in his article “Doubts about *cogito*” introduces several distinctions, namely

1. Knowing what is β vs. knowing that it is β ;
2. Awareness of a β vs. awareness that it is β ;
3. Direct vs. indirect awareness of facts;
4. An ability to distinguish between β s vs. an ability to tell they are β s;

He goes on to argue that awareness of what one thinks does not imply knowledge that one thinks. This is because the source of knowledge about what one thinks is quite distinct from the source of knowledge about that one thinks. Rejecting an “inner-sense” view of self-knowledge, Dretske argues that direct awareness of one’s thoughts (of what one is thinking), is never indicative of the fact that one is thinking. Hence, the source of one’s knowledge that one is thinking cannot lie in any special awareness of one’s thinking. Rather, it is something that one learns from others. Thus, Dretske concludes, it is plausible that at a young age children may know *what* they are thinking without knowing *that* they are thinking. Put differently, while we do have privileged access to the content of our thoughts, we need others to know that we are thinking. Dretske doesn’t tell us much about how we learn that we are thinking. However, his emphasis on the role of other persons already hints at the necessity of intersubjectivity (and possibly of being embedded in a linguistic community) for this learning process that is picked up and explored further in the following articles.

In her article “From consciousness to self-consciousness” Baker distinguishes sharply between consciousness on the one hand and self-consciousness on the other, and sets herself the task of sketching the developmental road from one to the other. While she concedes that there are nonconceptual forms of consciousness, which are characterized by a rudimentary first-person perspective and are shared with certain non-human animals, she argues – contra authors like Bermúdez (1998) – that there cannot be any nonconceptual forms of self-consciousness. This is because self-consciousness requires a robust first-person perspective, which, in turn, requires the possession of the linguistic first-person concept. Her distinction arguably resonates with Dretske’s claim that one can be aware of one’s thoughts without being aware that one is thinking – though Dretske is less explicit on the need for concept possession for

the latter. According to Baker, not only does self-consciousness require the ability to refer to oneself in thought or language by means of the first-person pronoun (as in the thought 'I am tall'), but it also requires the ability to attribute first-person reference to oneself (as in the thought 'I wish that I was tall'). Only the latter is evidence for a robust first-person perspective (and thus of the possession of a first-person concept). On Baker's view, concept possession requires the ability to apply the concept in question in a range of relevant circumstances and this ability, in turn, only emerges when one learns a language (and with it a range of empirical concepts that can then be applied to oneself). Thus, on this view, the ability to think of oneself as oneself and the ability to use the first-person pronoun are inextricably linked to each other.

Similar to Baker, Strasser claims in her paper "How minimal can self-consciousness be?" that self-consciousness requires the ability to ascribe properties to oneself (and thus to recognize oneself as the bearer of certain properties). However, she calls into question the view that the latter requires linguistic abilities. On her view, self-consciousness can come in various degrees, not all of which require language. Using the notions of the sense of ownership, authorship and agency she argues that we can distinguish three different senses in which a creature can be said to possess the ability to ascribe properties to themselves. Strasser's aim in the article is to investigate the cognitive abilities underlying each of these three senses, thereby developing criteria by which we can decide whether an infant can be said to be self-conscious or not. She concludes that minimal forms of self-consciousness can be present in the absence of linguistic abilities, whereas on a stronger notion of self-consciousness linguistic abilities are required. She then provides a detailed discussion of several phenomena from the developmental literature (including imitative behaviour, pretend play, perspective taking and theory of mind) in order to put the criteria she has established to work. To the extent that Strasser's stronger notion of self-consciousness is equivalent to Baker's notion of self-consciousness, Strasser and Baker seem to be in agreement. However, Strasser's notion – in contrast to Baker's – does not seem to require the ability to attribute self-reference to oneself. Moreover, Strasser's view makes room for a weaker notion of self-consciousness as well, which Baker would reject.

In her article "Self-consciousness and intersubjectivity" Musholt also makes the case for a conceptual distinction between consciousness on the one hand and self-consciousness on the other. Similar to Baker, she argues that it is one thing to have a perspective, but quite another to be aware of having a perspective. On her view, while the former implies the presence of implicit self-related information, the

latter requires explicit self-representation. And only the latter is indicative of self-consciousness. Musholt further argues that in order to account for the transition from self-related information that is implicit in perception and bodily experience to an explicit representation of oneself as a subject among other subjects one has to consider the role of intersubjectivity. Specifically, she claims that self-awareness requires an awareness of other subjects and the ability to contrast one's own mental and bodily states with those of others. Having alluded to the parallels of this approach with certain aspects of theories developed in the tradition of German idealism as well as in phenomenological approaches, she then uses insights from the cognitive sciences (in particular developmental psychology) to develop a multi-level account of this perspectival differentiation. Thus, her paper explicitly bridges the discussion about nonconceptual and conceptual forms of (self-)consciousness with the discussion of the role of intersubjectivity in acquiring self-consciousness.

The notion of intersubjectivity also plays an important role in Gallagher's paper on "The body in social context: Some qualifications on the 'warmth and intimacy' of bodily self-consciousness". Gallagher takes William James' discussion of the warmth and intimacy of bodily self-consciousness as his point of departure and discusses various pathologies associated with the awareness of one's body. The aim of the paper is, one the one hand, to demonstrate that the concepts of body image and body schema can be fruitfully applied to achieve a better understanding of such pathologies and, on the other hand, to show that bodily self-awareness is already permeated by social dimensions. The first part of the paper deals with a defence of the distinction between body image and body schema against a number of criticisms and against alternative approaches that focus on the neural correlates of the various aspects of (bodily) self-awareness. The second part of the paper introduces a novel aspect into the discussion – namely the role of intersubjectivity for basic, bodily self-awareness. Understanding the importance of the social dimension of bodily self-awareness is, according to Gallagher, also important for an understanding of various pathologies. Gallagher thus opens up a new and fascinating aspect of the debate on self-consciousness and its pathologies. It remains to be seen, however, how his notion of self-consciousness relates to the notions of self-consciousness developed in the preceding (and the following) papers. Another interesting question for future research that is raised by Gallagher's article is whether the evidence he cites should lead us to reject the principle of "methodological individualism" in favour of assuming an irreducibly intersubjective dimension of embodied social cognition.

Bridging the discussion of conceptual and nonconceptual ways of representing the self with the discussion of the temporal aspects of self-consciousness, Bermúdez' paper on "Memory judgments and immunity to error through misidentification" focuses on a feature that many authors take to be a central feature of self-conscious thoughts, namely their immunity to error through misidentification relative to the first-person pronoun. He argues that while first-person judgments that are made on the basis of autobiographical memories typically possess this feature, it can neither be accounted for in terms of the possession conditions of the relevant concepts involved in the self-attribution of the memories in question, nor in terms of their experiential features. Rather, Bermúdez claims, it is the fact that the remembered episodes are located within an autobiographical narrative that provides the key to an understanding of why a thinker who forms a first person judgement that is grounded in an autobiographical memory must ipso facto grasp that this judgement cannot misidentify.

The "narrative self" takes centre stage in some of the following articles as well.

In his paper "The time of the self", Zahavi distinguishes two philosophical conceptions of the self in its relation to temporality. On the first view, being a self (as opposed to merely being conscious or sentient) is an achievement, which depends on one's (more or less successful) self-understanding and self-interpretation. This, in turn, depends on the life story that we tell to ourselves and to others. It is only through the weaving of such a story that we can synthesize life's diverse aspects and experiences into a unique and stable identity. Importantly, as Zahavi points out, one's life story – which will be determined by one's values, goals and aims – cannot be understood in isolation from one's social community, as well as one's cultural and historical context. Thus, as Zahavi puts it, while we might be our lives' stories' main characters and narrators, we are not its sole authors. But, asks Zahavi, does the narrative account do full justice to the first-personal character of our conscious experience? On Zahavi's view, conscious experience is characterized by a specific mode of givenness – namely its first-personal character or "mineness". Importantly, this first-person character itself also possesses a temporal dimension. He argues that this experiential perspective of the subject is left out by narrative approaches. However, approaches that focus on the experiential perspective of the subject alone are also insufficient on his view, as they are unable to account for the persistence of the self across interruptions in the stream of consciousness. (Also see the papers by Dainton and Crone for a discussion of this problem.) He concludes that we need both accounts of the self: On the one hand, an account that emphasizes its experiential dimension, and in particular its first-personal character, which must include the temporality of this character. And on the other hand, an account that relies on the narrative approach in order to provide a dimension

of temporality and identity that extends beyond the horizon of first-personal givenness.

The intimate relation of temporality and self is also at issue in Dainton's paper "Selfhood and the Flow of Experience". Dainton's argument is based on the assumption that it is perfectly possible to fit a conception of a "self" into a naturalistic metaphysical framework. According to Dainton, the self is a subject of experience whose essential attribute is a persisting capacity for a unified and continuous experience. This view, however, seems to be challenged by two recent accounts on the self: Galen Strawson holds that selves are essentially conscious subjects who are nonetheless extremely short-lived (their duration may be less than a second). And Mark Johnston argues that selves are nothing but synchronically existing "arenas of presence". (A related view is also defended in Flanagan's paper in this issue.) Now, according to Dainton, there is no obligation to follow these two contributors in their rejection of a diachronically persisting self. He argues that neither Strawson nor Johnston do justice to an important phenomenological feature: the continuity of ordinary experience. If correctly spelled out, phenomenal continuity can be shown to be constitutive for selves: metaphorically speaking, brief phases of experiences directly flow into the next; they make up overlapping chains of diachronic co-conscious states. Crucially, according to Dainton, incidental disruptions of diachronic conscious experience are bridged by the subject's persisting capacity for continuous experiences. Selves can cease to be conscious and continue to exist provided they retain the potential to be continuously conscious.

Crone's article "Phenomenal Self-Identity Over Time" is directly linked to Dainton's argument as it criticises one central claim. She agrees with Dainton on the overall idea that experiential features of mental states are functionally related to the persistence of a person. However, as Crone argues, phenomenal continuity fails as a criterion for this persistence: streams of experience are interrupted every now and then (for instance, in dreamless sleep), and Dainton's strategy cannot properly solve this problem. (A similar criticism of Dainton's view is also expressed in Zahavi's paper.) As a consequence, Crone suggests to move away from the search for constitutive criteria of personal persistence and to instead turn to a different philosophical problem: to the analysis of a particular psychological state, namely the (non-inferential) awareness of self-identity over time. While the continuity of experience (the stream of consciousness) is doubtless an important source for a person's sense of self-identity over time, it cannot be the only one since it is unable to explain identity over longer

durations of time. Crone argues that it is certain experiential features of episodic autobiographical memories that underwrite our sense of self-identity over time. These ideas bear an interesting relation to Bermúdez' discussion of the immunity to error through misidentification relative to the first-person pronoun of judgements based on episodic autobiographical memories.

The distinction between different philosophical conceptions of the self and its relation to time that is drawn by Zahavi is echoed in a distinction urged by Flanagan in his paper "Phenomenal and historical selves". Flanagan argues that on one view of personhood, which he associates with John Locke, and which he terms the forensic-narrative view, a person is seen in abstraction from self-experience. This abstraction is based on autobiographical memory and is related to the narrative accounts described by Zahavi. According to Flanagan, it can be seen to be useful for the way in which we navigate our social sphere, in particular in the context of holding each other responsible for past actions. However, Flanagan claims, this view also necessarily remains impoverished and restrictive and we should consider replacing it with a more basic and more flexible view of how different individuals experience themselves as persons, which he associates with William James and Galen Strawson. Flanagan argues that while we remain continuous historical beings throughout our lives, we can experience a disconnect between the way we phenomenally experience ourselves in the present and the kind of person we were in the past (or will be in the future). Our 'streams of consciousness' can assume different patterns over time, as it were. Thus, he seems to be in agreement with Zahavi's claim that narrative accounts cannot do full justice to the experiential dimension of being a conscious self. However, while Zahavi, Crone and Dainton emphasize the unifying nature of the first-person perspective, Flanagan asks us to consider the possibility of experiencing ourselves as rather different from the person we used to be. If he is right, we might in a certain sense at different points in our lives literally become different persons – there might simply not be a unique self-identity that is preserved throughout our lives. Acknowledging this, according to Flanagan, not only gives us a philosophically richer and more accurate view of personhood, it is also frees us up to recognize our experiences for what they are and to engage more authentically in the project of self-understanding and self-improving.

As this brief overview shows, the articles in this special issue address and bridge many different facets of and approaches to the study self-consciousness. We hope that this issue demonstrates the value in seeking a more integrated approach in the study of

self-consciousness and that it will provide the point of departure for new research and debate in this area.

The inspiration for this special issue first grew out of a workshop on “Facets of self-consciousness” that was organized by the three editors and held at the Berlin School of Mind and Brain in March 2010. We would like to thank all the presenters and participants in this workshop for their stimulating contributions, and the School and the German Research Foundation (DFG) for their support.

References

Bermúdez, José L. 1998: *The Paradox of Self-Consciousness*. Cambridge: MA: MIT Press.

Bieri, Peter 1986: "Zeiterfahrung und Personalität". In: Heinz Burger (ed.), *Zeit, Natur und Mensch*. Verlag A. Spitz, 261-281.

Crone, Katja 2009: "Selbstbewusstsein und Identität - die Funktion der qualitativen Erlebnisperspektive". In: Jan Heilinger, Matthias Jung (eds.), *Funktionen des Erlebens*. Berlin/New York: de Gruyter, 337-363.

Dainton, Barry 2000: *Stream of Consciousness. Unity and Continuity in Conscious Experience*. Routledge.

Evans, Gareth 1982: *The Varieties of Reference*. Oxford: Oxford University Press.

Frank, Manfred 1991: *Selbstbewusstseinstheorien von Fichte bis Sartre*. Frankfurt/Main: Suhrkamp.

Gallagher, Shaun 2005: *How the Body Shapes the Mind*. Cambridge University Press.

Gallagher, Shaun 2007: "Pathologies in Narrative Structures". In: Dan Hutto (ed.), *Narrative and Understanding Persons*. Cambridge University Press, 203-224.

Gallagher, Shaun and Hutto, Dan 2008: "Understanding others through primary interaction and narrative practice". *The shared Mind: Perspectives on intersubjectivity*, 17-38.

Happé, Francesca 2003: "Theory of mind and the self". *Annals of the New York Academy of Sciences* 1001, 134-144.

Hurley, Susan 1997: "Nonconceptual self-consciousness and agency: perspective and access". *Communication and Cognition*.

Husserl, Edmund 1973: *Zur Phänomenologie der Intersubjektivität. Texte aus dem Nachlass. Teil I -III (Husserliana 13-15)*. Den Haag: Martinus Nijhoff.

Husserl, Edmund 1985: Texte zur Phänomenologie des inneren Zeitbewußtseins. Hamburg: Meiner.

Legrand, Dorothée 2006: "The bodily self: The sensori-motor roots of pre-reflective self-consciousness". *Phenomenology and the Cognitive Sciences* 5, no. 1, 89–118.

Legrand, Dorothée 2007: "Pre-reflective self-as-subject from experiential and empirical perspectives". *Consciousness and Cognition* 16, no. 3, 583–599.

Schechtman, Marya 1996: *The Constitution of Selves*. Cornell University Press.

Spitzley, Thomas 2000: *Facetten des Ich*. Paderborn: Mentis.

Strawson, Peter F. 1959: *Individuals: An essay in descriptive metaphysics*. Routledge.

Vosgerau, Gottfried 2009: *Mental representation and self-consciousness: From Basic Self-Representation to Self-Related Cognition*. Paderborn: Mentis.

Wellman Henry M. et al., 2001: "Meta-Analysis of Theory-of-Mind-Development: The truth about false belief". *Child Development* 72, 665-684.

Zahavi, Dan 2003: "Inner Time Consciousness and Prereflective Self-Awareness". In: D. Welton (ed.), *The New Husserl. A Critical Reader*. Indiana University Press, 157-180.

Zahavi, Dan 2005: *Subjectivity and selfhood: Investigating the first-person perspective*. MIT Press.

Zahavi, Dan 2007: "Perception of Duration Presupposes Duration of Perception – or Does it? Husserl and Dainton on Time". *International Journal of Philosophical Studies* 15 (3), 453-471.