Recent German books in the philosophy of mind

Michael Esfeld & Michael Sollberger
University of Lausanne, Department of Philosophy
CH-1015 Lausanne, Switzerland
(published in *Dialectica* 62(1) (2008), pp. 128–135)

Bartels, Andreas (2005): *Strukturale Repräsentation*. Paderborn: Mentis.
Beyer, Christian (2006): *Subjektivität, Intersubjektivität, Personalität. Ein Beitrag zur Philosophie der Person*. Berlin: de Gruyter.
Gottschling, Verena (2003): *Bilder im Geiste. Die Imagery-Debatte*. Paderborn: Mentis.
Nida-Rümelin, Martine (2006): *Der Blick von Innen. Zur transtemporalen Identität bewusstseinsfähiger Wesen*. Frankfurt (Main): Suhrkamp.
Pauen, Michael (2004): *Illusion Freiheit? Mögliche und unmögliche Konsequenzen der Hirnforschung*. Frankfurt (Main): Fischer.
Walde, Bettina (2006): *Willensfreiheit und Hirnforschung. Das Freiheitsmodell des epistemischen Libertarismus*. Paderborn: Mentis.
Willaschek, Marcus (2003): *Der mentale Zugang zur Welt. Realismus, Skeptizismus und Intentionalität*. Frankfurt (Main): Klostermann.

Andreas Bartels argues in his book for a structuralist theory of representation. He sets three targets that such a theory should meet: (a) it should explain what turns something into a representation of something else; (b) it should give an account of why a representation provides information about something else; and (c) it should leave room for misrepresentation (17). Bartels conceives the structuralist theory of representation as providing an answer to (b) in the first place. He sets out a structuralist account of the content of representations combined with a functional view of (a) and a causal theory of the reference of a representation. Distinguishing between the account of content and the account of reference in this way is supposed to enable the resulting position to deal with misrepresentation (53). A structuralist theory of representational content links up with the tradition of image theories of representation but is distinct from them, not falling victim to the objections that have discredited that tradition. The representational content consists in an isomorphism between the structure of one entity, the representation, and the structure of another entity, usually the target of the representation. Bartels makes this idea precise in chapter 2.1. He then applies this conception to a broad field, starting with fiction (chapter 3) and moving on from there to representation in scientific models (chapter 4). In particular, Bartels discusses space-time models and the view that what these models represent, space-time, is itself a structure (chapter 4.4). The main part of the applications in what follows concern mental representation. Bartels considers conceptual content (chapter 5), qualia (chapter 6) and representational theories of behaviour (chapter 7). He finally points out that his account is compatible with but not tied to realism (chapter 8).

The strong side of the book is that it offers a structuralist theory of the content of representations in contrast to mainstream causal and functional theories and shows how this theory can be applied to a wide range of fields. Of course, the drawback of such a wide range is that for each of these fields, one may ask for further details. Consider the most illustrative example in that respect, qualia: Bartels convincingly points out shortcomings in Dretske's representationalist and externalist theory of qualia (133-150). Although he certainly has a point in claiming that his theory fares better than Dretske's, there remain of course a lot of objections to

be countered in order to establish a representationalist theory of qualia. Nonetheless, Bartels' book is an important contribution to the debate, setting out a new position with great potential and thus providing a perspective for further fruitful work.

The overall aim of Verena Gottschling's book, which is based on her PhD Thesis, is to make clear what is really at stake in the so-called *imagery debate*. To do this, she sets out the tenets of *pictorialism* and *descriptionalism*, discusses them in the light of the most influential objections and finally wonders whether a pictorialist conception of mental imagery can be intelligibly championed. Her thesis is that the equivocal use of the notion of picture-like representations is at the bottom of many conceptual confusions. Her book identifies these shortcomings and, after displaying the variety of distinct pictorialist positions, aims to show that the strong "Real Picture" reading is the only promising candidate.

Gottschling points out that the basic assumption of cognitive science, i.e. the computational representational theory of thought (CRTT), provides the vital framework within which the debate takes place (e.g. §2.1, §5.2). Basically, pictorialists, whose most prominent figure is Stephen M. Kosslyn, have to show two things: First, the underlying representational contents of mental images are depicted in a spatial medium, and second, the spatial properties of these imagistic representations play a causal role in the subject's cognitive architecture (chapter 2). Gottschling goes on to examine whether there are common properties shared by physical pictures and mental images that can justify the talk of mental, *picture-like* representations (chapter 4). Finally she emphasizes that pictorialism splits into two diverse strands: the weak "functional thesis" (FT) claims that mental images only function like pictures, whereas Kosslyn's strong "Real Picture thesis" (RPT) identifies images with the topographic maps in the cortex (159). That is, RPT claims that mental images are real two dimensional displays in the human brain (242).

Chapter five deals with Zenon Pylyshyn's descriptionalist arguments in favour of tacit knowledge (§5.3.1) and cognitive penetrability (§5.3.3). If valid, they yield the conclusions that (a) no depictive representations are needed for accommodating empirical data of imagery, and that (b) CRTT is incompatible with RPT. The rest of the book is devoted to spelling out the positions of FT and RPT. On the one hand, it remains unclear what functional picture-like representations could be; it is therefore impossible to evaluate FT's explanatory force. On the other, Kosslyn's reductionist RPT is committed to abandoning CRTT.

Unfortunately, Gottschling's examination does not go beyond this conceptual disentanglement of the imagery debate. For instance, it would have been interesting to delve into an elaborated theory of RPT in order to show how the picture-like properties of representation can play a causal role in the subject's cognition. Nonetheless, Gottschling's books sets out the conceptual framework within which a future, fully-fledged theory of pictorialism should be developed.

Marcus Willaschek's book, based on his Habilitation Thesis, is a defence of direct realism as regards everyday medium-sized empirical objects. He develops a radically externalist conception of the mind in order to make intelligible how empirical thoughts can be directly about a thought-independent reality. Right from the beginning, Willaschek suggests that the whole discussion about realism rests on a misunderstanding, for the point at issue is not the world's independence but how we can be cognitively related to it.

Willaschek emphasizes that everyday common sense is committed to a maximally realist position, comprising both a semantic and an epistemic claim: our empirical beliefs are directly about a thought-independent reality (46) and we possess knowledge about this reality (47). Willaschek identifies three lines of reasoning that make direct realism look dubious: Cartesian

scepticism and its resulting internalistic representational theory of the mind (§§ 22-25), ontological relativism that conceives the mind-world relation as always being mediated by symbolic forms (§§ 26-29), and empiricist and meaning theoretic forms of verificationism (§§30-32).

Having rebuffed scepticism in chapter 4, Willaschek then sets out his own normative-disjunctive account of the relation between mind and world (chapter 5). According to his disjunctive token-externalism, the content of empirical beliefs is not neutral with respect to their truth-value – they are fact-dependent. A belief can be about the fact that p in two different ways, namely in being true because of p and in being false because of non-p, and this is why beliefs are directly about the thought-independent world (223). Even if sometimes subjectively indistinguishable, true and false beliefs therefore belong to different metaphysical kinds. Willaschek goes on to show that the intrinsically normative dimension of beliefs, i.e. that conceptual content is sensitive to truth and justification, can be accommodated by a disjunctive account of perceptual experience (§60). Perceptual content is metaphysically transparent and, as Kant argued, inextricably made up of both phenomenal and conceptual elements (275). This passive articulation of concepts in perception can play the role of justification in our everyday epistemic practice, since the world as it is in itself guides our beliefs.

Despite being very clear and precise Willaschek's line of thought may give rise to two worries. First, one might be sceptical about the confidence Willaschek has in the common sense of Western Europeans: Why should we unreservedly subscribe to the authoritativeness of everyday common sense vis-à-vis philosophical thought? After all, this form of common sense might be nothing but prejudice. Second, since he assimilates his normative-disjunctive account of intentionality to the disjunctivist model of perception, one would have wished Willaschek to engage more deeply with disjunctivism and the various challenges it faces. Nevertheless, this book provides a concise overview of the realism debate and develops an original theory of the mind-world relation.

Martine Nida-Rümelin's book, based on her Habilitation Thesis, is an elaboration of the view that there are no underdetermined cases of transtemporal personal identity. Even if Andrea's brain undergoes an operation such that the left half of her brain wakes up in one body and the right half in another body, there is a fact of the matter consisting in exactly one of the following four outcomes: Andrea is identical either with left-Andrea or with right-Andrea or with none of them or with both of them (being one person that has two bodies). The same goes for Derek whose body is destroyed on Earth and reproduced twice one two different planets by beaming technology. Nida-Rümelin bases this view exclusively on the intuitions we have regarding ourselves. She explains these intuitions (introduction, chapters 1 and 2), enquires into their conceptual roots (chapter 3) and considers the metaphysics of the self based on them (chapters 4 and 5).

Nida-Rümelin contrasts this realist view with reductionism, that is, the claim that facts of personal identity reduce to empirical or psychological facts. Personal identity requires no criteria (e.g. 35-36, 171-202). The fact that the two copies of Derek are identical in all empirical and psychological respects at the moment they come into existence does not prevent one of them rather than the other one being identical with Derek. Thus, the fact of being this person is a primitive fact. Such a view seems to lead to substance dualism. Nida-Rümelin actually pleads in favour of taking substance dualism seriously (318-324) but she considers her realism with respect to transtemporal personal identity to be compatible with materialism as well (298, 311-313). She even argues that the realist in her sense should commit herself to a supervenience thesis, namely to strong nomological supervenience (chapter 4).

However, it is not evident that strong supervenience (even if applied only to all nomologically and not to all metaphysically possible worlds) is compatible with substance dualism and the claim that transtemporal personal identity is independent of any criteria. One might argue that if there is strong nomological supervenience, then in all nomologically possible worlds there are sufficient empirical conditions for transtemporal identity. Consequently it seems that if we knew those conditions, we could reduce the description of cases of transtemporal personal identity to the description of such conditions. When talking about supervenience, Nida-Rümelin admits that there may be sufficient empirical conditions for transtemporal identity (e.g. 243, 280-287), while at other places she envisages cases that are conceived as empirically indiscernible and yet differ with respect to transtemporal personal identity (e.g. 27-28, 45, 213, 240-245, 313, 340).

Nida-Rümelin's book is a detailed, thorough and clear defence of a particular view of personal identity that poses a serious challenge to the mainstream accounts. Nonetheless, one may voice the following queries: Are our intuitions concerning the transtemporal identity of persons sufficient to ground her view of transtemporal identity? Even if this were so, would it be appropriate to base a metaphysics of the self only on these intuitions, whatever the consequences for other intuitions about the mind (e.g. mental causation) and knowledge about ourselves stemming from other sources (e.g. scientific knowledge) might be?

Christian Beyer also considers personal identity in his book, which is based on his Habilitation Thesis. The book is intended to sketch out a comprehensive theory of what it is to be a person and is based on analytic philosophy and Husserl's phenomenology. Beyer focuses on the interplay between subjectivity and intersubjectivity. In the first part of the book, he sets out a version of the higher-order theory of intentional consciousness, namely a theory of indexical metabeliefs. A person is in a conscious intentional state if and only if she non-inferentially believes that she is in the state of the sort in question (25-26). The metabelief in question is a dispositional state (chapter 1). Beyer sets out in what manner this conception sheds light on the synchronic and the diachronic unity of consciousness (chapter 2).

The main – and controversial – claims of the second and the third part concern the importance of intersubjectivity for the conception we have of ourselves and the world. Beyer supports the following four hypotheses: (a) Only intersubjective experience in attributing intentional states to other persons justifies a person in believing in her own identity as a subject who experiences different intentional states. In this context, Beyer pleads in favour of a moderate version of the simulation theory of attributing intentional states to other persons (chapters 3 and 4). (b) Only the mutual attribution of intentional states in (radical) interpretation under the aspect of what Davidson calls "triangulation" justifies our belief in an objective, spatiotemporal world (chapter 4). (c) Only certain conventional, intersubjective criteria of physical or psychological continuity justify our belief in the transtemporal, numerical identity of a person (chapter 5). (d) Only the empathic experience of metavolitions in another person justifies the belief that the other person is a subject in the moral sense (chapter 6).

These are claims about the conscious *justification* from a first-person point of view of certain beliefs that are central to our view of ourselves; the justification of these beliefs depends on intersubjective experience. Beyer does not go so far as to maintain that intersubjective experience is a necessary condition for the *existence* of the objects of those beliefs, e.g. that intersubjective experience is a necessary condition for personal identity. The book leaves these metaphysical issues open. It is clearly and concisely argued and covers all the central aspects of what it is to be a human person from the perspective of the importance of intersubjective experience for the conception that we have of ourselves.

In his book about free will, Michael Pauen advances two major claims: First, an action or act of will is free if and only if it is self-determined. Second, Pauen argues against the coherence of the incompatibilist idea according to which a *person* could initiate a causal chain. The new feature of his so-called *minimal conception of personal freedom* (59) is the development of a kind of compatibilism based only on the uncontroversial assumption that freely willed actions are events that neither happen by chance nor by coercion. He concludes from this assumption to the principles of autonomy and authorship (63): acting of her own free will, an agent's action not only originates in (autonomy) but also emanates from (authorship) the agent to whom it can thus be intelligibly ascribed. Pauen thinks of the self in psychological terms in that specific traits like desires, wants and beliefs constitute a person's self. Also called personal preferences (71), these specific traits make it intelligible why agent A did action X and not action Y and, therefore, set up a robust relation between A and her action. So if the action is determined by A's personal preferences, then it is self-determined and hence freely willed (163).

After having laid out his positive account, Pauen seeks to refute incompatibilism. He claims that both the strong interpretation of the principle of alternative possibilities (106) and the consequence argument (136) rely upon the incoherent assumption that freedom of will calls for an agent who is a completely undetermined starting point of a causal chain. Basically, Pauen thinks that incompatibilism infringes upon the principle of autonomy: Given that being a person requires one to have personal preferences, it is inconceivable that one can link A qua *person* with her alleged action if A's decision to do X instead of Y were completely undetermined. After all, X depends on A's personal preferences that, in turn, are central determinants of A.

Even if *prima facie* highly plausible one might wonder whether Pauen's compatibilist scheme of self-determination is sufficient for free will. It intuitively seems that the agent's personal preferences can arise in a freedom-undermining way, as when they are the mere product of education, genes, hormones, etc. Pauen's overall line of thought is, however, concise and articulated in a fresh, non-technical style.

Bettina Walde argues for three theses in her book on free will, which is based on her Habilitation Thesis:

(a) Mental state tokens belong to the physical realm, that is, are identical with brain state tokens. Otherwise, they would not be causally efficacious, but epiphenomenal.

(b) Mental intentions to act (the will) are determined in an appropriate causal manner.

(c) The future is open from the perspective of the agent.

The agent has to be in a position such that considering different future courses of action is meaningful to her. Taken together, these three conditions are sufficient for free will. Walde bases the first condition on the well-known argument from the causal completeness of physics (chapters 3 & 10). She uses this condition in order to argue that those neuroscientists who claim to have discovered empirical evidence against free will implicitly presuppose dualism (16-20). For if mental intentions are identical with brain states, epiphenomenalism with respect to mental intentions implies that there are some brain states that are epiphenomenal; but there is no such empirical evidence. The second condition takes up compatibilism and is intended to avoid the objection that if mental intentions had no causes, they would be indistinguishable from chance events (chapter 11). The third condition pays tribute to libertarianism and the consequence argument. It is why Walde calls her position "epistemic libertarianism". However, the libertarianism is only epistemic. The world may be deterministic and an intelligent being outside of the universe may be able to predict everything that happens in the universe (21, 132, 158). It is only required that the agent cannot know her own future (chapter 12).

Consequently, the metaphysics that Walde defends clearly falls within compatibilism. It seems therefore more appropriate to characterize her position as a form of compatibilism that poses the openness of the future from the agent's perspective as one of the conditions under which there is free will in a deterministic world. In other words, acting on the basis of considering several future courses of action to be open counts among the conditions that determine a will that is free (cf. 176-178, 187-188, 193, 200-206). Walde's overall argument is precise, well structured and convincing, refuting the recent objections by neuroscientists to free will and setting out a position that takes into account both the prephilosophical view of ourselves and our knowledge of the world, including our brain.

This selection of books shows that German philosophy of mind is up to date, covers a wide range of topics that are hotly debated today. It is evident that there is a strong interest in taking into account empirical findings stemming from the cognitive sciences. Important contributions to the international debate are thus to be expected from German philosophy of mind.