Erasmus Student Journal of Philosophy

#1 | December 2011

ESJP
#1 | 2011

Editorial

There is something intriguingly exciting about the formative years in the lives of great philosophers. For it is here we find their first encounters with the intellectual tradition of which they would later become a part, the questions that triggered their thinking, and the moment that their ideas began to take on distinctive shapes of their own. The Erasmus Student Journal of Philosophy (ESJP) has been founded to capture and share the philosophical excitement in these formative years of the most talented students at the Faculty of Philosophy of the Erasmus University Rotterdam.

Having the honour to present this first issue of the ESJP, I am filled with pride and a deep sense of gratitude toward everyone who has contributed to make this possible. It has been a privilege to see the humble idea that the best work of our students deserves a podium evolve into the fruitful cooperation between students and teachers that is now the ESJP. Hopefully the ESJP will become an institution of our faculty that continues to enrich the philosophical environment in which students develop their thinking for many years to come.

There are numerous people without whom this first issue of the ESJP would not have been possible. I would like to express my deep-felt gratitude to dr. Gijs van Oenen, dr. Awee Prins and prof. dr. Ingrid Robeyns for their encouragement and advice when the ESJP was nothing more than an idea, roughly outlined in a five page proposal; prof. dr. Wiep van Bunge, prof. dr. Jos de Mul, prof. dr. Ingrid Robeyns and prof. dr. Jack Vromen for endorsing and approving that proposal; prof. dr. Wiep van Bunge, dr. Patrick Delaere, and dr. Fred Muller for their willingness to keep an eye on its further evolution; the lecturers who nominated the best essays that were written for their courses; our anonymous referees for their excellent comments and advice; Janneke Boerman for her legal advice concerning copyright; Amanda Koopman for her help with the website and communications; Ivo Jeukens for the elegant design of our layout; the contributing authors for their essays and their excellent and positive responses to the feedback of our referees; and of course our editors Thijs Heijmeskamp, Julien Kloeg, Myrthe van Nus, and Volker Ruitinga, not only for their dedication and outstanding work, but also the cooperation that has made this whole undertaking into a joyful experience. I sincerely hope that in return for all the efforts that have come together in this first issue of the Erasmus Student Journal of Philosophy, pride and satisfaction about the end-result will now flow back to everyone mentioned above.

Daan Gijsbertse

Editor-in-Chief

ESJP
#1 | 2011

About the Erasmus Student Journal of Philosophy

The Erasmus Student Journal of Philosophy (ESJP) is a double-blind peer-reviewed student journal that publishes the best philosophical papers written by students from the Faculty of Philosophy, Erasmus University Rotterdam. Its aims are to further enrich the philosophical environment in which Rotterdam's philosophy students develop their thinking and to bring their best work to the attention of a wider intellectual audience. A new issue of the ESJP will appear on our website (see below) every July and December.

To offer the highest possible quality for a student journal, the ESJP only accepts papers that (a) have been written for a course that is part of the Faculty of Philosophy's curriculum and (b) nominated for publication in the ESJP by the teacher of that course. In addition, each paper that is published in the ESJP is first subjected to a double-blind peer review process in which at least one other teacher and two student editors act as referees.

The ESJP highly encourages students to write their papers for courses at our faculty with the goals of publishing in our journal and appealing to a wider intellectual audience in mind.

More information about the ESJP can be found on our website:

www.eur.nl/fw/esjp

Contact

esjp@fwb.eur.nl

Disclaimer

ESJP
#1 | 2011

In this issue

If there is a common theme to the contributions in this first issue of the ESJP, it is the original and critical way in which the authors take on the work of established philosophers and schools of thought in different fields of study.

In *The Foreground and Background of Consciousness*, Mara van der Lugt questions accounts of the self commonly arrived at by philosophers through introspection, especially within the context of the free will debate. Introducing an alternative mode of introspection, she argues that these accounts are at odds with how we normally experience ourselves and proposes to introduce a foreground and background to our understanding of the experience of consciousness.

Elize de Mul offers a refreshingly original perspective on the most prominent theories in the history of hermeneutics. Based on the game theories of Johan Huizinga and especially Roger Caillois, she reinterprets the theories of interpretation by Schleiermacher, Dilthey, Gadamer, and Derrida as different forms of play within a larger game of meaning. Her essay *Het spel der betekenissen* is written in Dutch.

In *The Argument for Anomalous Monism, Again*, Deren Olgun demonstrates how even a mature philosophical debate can veer off track when participants fail to fully understand their opponent's philosophical position. With an impressive display of analytical rigor, he exposes an ontological misconception behind attacks on Davidson's position of Anomalous Monism in the mental causation debate.

Karin de Bruijn uses the image of 9/11's 'falling man' and Sophocles' Antigone, to reveal the tragedy in the crises of our time. Inspired by Judith Butler and Slavoj Žižek, she calls for an aestheticisation of mourning in order to come up with strategies of moderation and reconciliation to counter these crises. Her essay *Geschreven met licht* is also written in Dutch.

Last but not least, Volker Ruitinga puts John Rawls´ distinction between Ideal and Non-Ideal Theory, which so often figures in contemporary debates within social and political philosophy, into a broader perspective. In *Ideal Theory and Utopia* he calls on the literature about Utopian Theory to redefine the Rawlsian notion of Ideal Theory as a Utopia of social justice.

ESJP
#1 | 2011

Table of contents

ESJP
#1 | 2011

The Foreground and Background of Consciousness – An Introspective Argument Against Introspection

Mara van der Lugt

As for the pure philosophical 'freedom of the will' my will is as free as I feel it to be and there is an end to the matter.

– R.B. Braithwaite.[1]

There is a widespread tendency in the philosophy of our times to 'get back to basics'. In the wake of philosophical reconstruction workers such as G.E. Moore and Ludwig Wittgenstein, and the ordinary language philosophy they inspired, contemporary philosophers appear to have learned some important lessons: to take into account our most commonsensical notions about ourselves, each other, and the world we live in, and to beware of framing elaborate problems in theory where practically there are none. The result of such a common sense mentality is not only that our ordinary phenomenological apprehensions of life in general are taken seriously – as they should be, if philosophy is to make sense of the world we live in – but also that certain assumptions about our self-experience are allowed to go unquestioned. Thus we see that in many philosophical texts, especially those discussing free will, consciousness, agency, and deliberative awareness, a specific phenomenological[2] account of the self is presupposed: namely, the self as a fixed entity, a unified centre of consciousness, which focuses its intentions into actions, and hence experiences itself as the source of its actions. Such conceptions are usually based on philosophers' armchair introspection, yet they are deemed to be universal among mankind, an undeniable part of our most basic experience and common sense. Indeed, what can be more commonsensical than that we experience ourselves as selves?

I intend to question this assumption, particularly in the context of the free will debate. I do believe that such a perception of the self may arise when we put on our 'introspectacles': that is, when we are consciously thinking or talking about ourselves, and when we philosophically focus on 'the self itself'. However, I do not believe that it is warranted on the mere basis of such introspection to infer that this is *how we usually experience ourselves*. Could it not be the case that the often-reported experience of the 'self' and some of its attributes is a *result* of focused introspection, rather than an expression of our default self-experience? Perhaps the experienced 'self' *prior* to introspection encompasses a wider and more dynamic background of conscious, semi-conscious and preconscious aspects, while introspection focuses on what is at the foreground of our minds, and translates this into the experience of a consciously willing self.

This paper will focus on the nature of the introspective move itself, on the interpretation that may be implicit in such a moment of 'looking into oneself'. The main question to be answered is as follows: is the 'self' we encounter in introspection phenomenologically identical to how we *usually* experience ourselves? In what follows I will first of all survey the predominance of this 'introspective self' in the free will debate, and briefly discuss some previous criticisms of philosophers' use of phenomenology, in order to prepare a deeper critique of introspection. Second, I will try to expose several problems stemming from the nature of introspection itself, and argue that it is at least *possible* for the 'introspective self' to differ from the default 'experiential self'. Third, I will – perhaps paradoxically – call on introspection for evidence that such a divergence between introspected and experienced selves is phenomenologically plausible, since a different kind of introspection may lead to different introspective results. To eluci-

date this 'introspective argument against introspection' I will introduce the concepts of a foreground and background of the experience of consciousness. Finally, I will conclude that it is time to further problematise the use of introspection in philosophy, and briefly consider the implications of these arguments for the free will debate.

1. The 'Phenomenal Fact'

While there is no scarcity of phenomenological universals[3] concerning the nature of the self in any type of philosophy, it is in the context of the free will debate that they make their most prominent appearance. The free will problem can be construed as 'an unexplained gap between the category of physical phenomena and the category of subjective phenomena' (Libet, 1999: 55) – that is, as a discrepancy between the implausibility of mental causation (i.e. that our mental states can cause physical events – see Hohwy below) from a metaphysical point of view, on the one hand, and our experience of ourselves as conscious agents, on the other.[4] To put it more bluntly: philosophy may tell us we are not free, but 'we' feel we are. Solutions to the problem usually consist of either an attempt to bridge the metaphysical-experiential gap, or an effort to show that one of the two conflicting terms should be prioritised. Either way, the notion of a unitary, consciously acting self is invoked by many proponents of mental causation in order to voice an innermost experience, to which we all are supposed to be committed. At the same time, this notion is taken seriously by opponents of mental causation, who likewise consider it a basic part of our experience, be it one that must be explained away. Both sides of the debate, then, broadly seem to agree that the introspective experience of the acting self is a hard 'phenomenal fact' (to borrow a phrase from Libet, below), yet they differ in the value they assign to it: whether as a piece of '*prima facie* evidence', or as an illusion that should not be given philosophical credence.[5]

Consider the appearance of this 'phenomenal fact' in a variety of papers (I have highlighted the universalising tendencies in bold):

> '[...] we must recognize that **the almost universal experience** that we can act with a free, independent choice provides a kind of *prima facie* evidence that conscious mental processes can causatively control some brain processes [...] The phenomenal fact is that **most of us feel** that we do have free will [...]' (Libet, 1999: 56).

> '[...] the agency theory is appealing because it captures **the way we experience** our own activity. It does not seem to me (at least ordinarily) that I am caused to act by the reasons which favour doing so; it seems to be the case, rather, that I produce my own decisions in view of those reasons [...]' (O'Conner, 1995: 196).

> '[...] **your phenomenology** presents your own behavior to you as having yourself as its source, rather than (say) presenting your own behaviour to you as having your own occurrent mental events as its source [...]' (Horgan et al., quoted in Nahmias *et al.*, 2004: 167).

> '**Many people, including most philosophers,** have a very firm belief that there is mental causation, that is, that mental states such as beliefs, desires, intentions, and emotions are efficacious [...] in the causing of some physical events such as bodily movements and actions in the wider sense [...] **We thus have a very deep attachment** to mental causation' (Hohwy, 2004: 377).

> '[...] it seems **to each of us** that **we** have conscious will. It seems **we** have selves. It seems **we** have minds. It seems we are agents. It seems **we** cause what **we** do' (Wegner, 2002: 342).

> 'Human freedom is just **a fact of experience**. [...] a series of powerful arguments based on facts of **our own experience** inclines **us** to the conclusion that there must be some freedom of the will because **we all experience it** all the time' (Searle, 1984: 88).

What emerges from an overview of such 'low-brow' accounts of phenomenology – and many other examples can be found – is an informal

yet introspectively plausible sketch of the thinking, willing, acting self. The 'actish phenomenal quality' (Ginet, 1990: 13) at the source of these introspective musings springs from individual philosophers, yet their conclusions are universalised in a subtle move from 'I' to 'we', 'each of us' or 'most of us'.

Such practices of introspection or first-person phenomenology have been subjected to a variety of criticisms in the past decades, which in turn have given rise to new methods and versions of phenomenological investigations. For instance, Eddy Nahmias *et al*. have criticised the 'universality assumption' of free will philosophers who believe their own (often mutually incompatible) introspective reports are indicative of the human condition in general (2004: 164). Instead, the authors propose 'systematic psychological research on the relevant experiences of non-philosophers' (169) – that is, they attempt to gain insight to lay phenomenology through a series of empirical queries, and conclude that the 'universality assumption' is often wrong. Philosophers usually 'introspect through the lens of their theoretical commitments' and are therefore 'the wrong subjects to trust' (163).

Daniel Dennett appears to be expressing a similar concern in his critique of 'the *first-person-plural presumption*': when thinking about consciousness, 'I' is often widened into 'we' (1991: 67). Not only is this presumption unwarranted, says Dennett, but our very notion of infallible introspection is a mistake: 'I suspect that when we claim to be just using our powers of inner *observation*, we are always actually engaging in a sort of impromptu *theorizing* [...]' (*Ibid.*). There is no way of eliminating interpretation from such observation – thus all (not just philosophical) introspection is suspect. Yet Dennett does believe it possible to construct an objective method of phenomenology, which avoids the temptations of first-person introspection without sliding into a fully reductive behaviourism that forbids any talk of mental events. This 'heterophenomenology' describes subjects' reports in the scientific third-person perspective, and treats the reported intentional objects like fictional entities, which may or may not be real (*Ibid.*: 71-98).

These alternative methods of phenomenology, and others of the sort, share common ground in that they question the salience of individual introspection and tend towards a broader scientific survey of human experience. Herein lies an important insight. Yet these methods do nothing to answer my question about what happens in the introspective move: they criticise phenomenological claims just with respect to their *universalisation* of introspection, not with respect to introspection itself. The more fundamental question about introspection, I would argue, cannot be evaded, even by Nahmias and Dennett. If we want to avoid radical behaviourism and be able to say anything about consciousness or experience, as both authors do, somewhere along the line *someone* has to introspect – whether it be the philosopher or the subjects of an experiment. The Nahmias queries, though distanced from philosophers' introspection, remain dependent on the lay subjects' introspection. Dennett's heterophenomenologist is merely re-describing the first-person experiences of the subject into a third-person theory. (The scientist who notes that 'subject S reports having an experience E' may not have to introspect personally, yet still relies on S introspecting and commenting on E.) Either way, empirical queries of phenomenology cannot entirely steer clear of the first-person perspective. Hence, the 'deeper' question concerning introspection has not yet been answered. On the contrary – it has not even been posed.

2. The Phenomenological Fallacy

The deeper problem, remember, is whether introspection adequately represents our common experience of ourselves to ourselves. To argue that such an assumption of adequacy is at least mildly suspicious and perhaps even highly problematic, I will make two remarks – one might call them premises – on which to build my case. Since my argument is designed to make a general point, it should not depend on a specific theory of either phenomenology or introspection, and I do not want to commit to one. I will therefore not introduce detailed theoretical and methodological underpinnings of the topics in question, and merely outline some points that should be *relatively* modest and uncontroversial.

The first point should be clear from the discussion above: phenomenology is not like most 'normal' sciences. It has for its subject matter not the outer world as it is, but as it appears to the inner world. As the case of

Dennett shows, even if we aspire to objectively describe the phenomenon, in phenomenology we are ultimately bound to some kind of subjective experience: to the first-person perspective. If phenomenology wants to investigate consciousness and experience – what it is like to be or see or feel something – it has to rely on the 'inner take' on things. It cannot naively fall back on the scientific 'outer take' without forfeiting the project itself, as is the common tendency in behaviourism. Phenomenology without a *view* from the inside, I would argue, is not phenomenology at all.

The basic subjective commitment of phenomenology opens a realm of possibilities as well as problems of its own. For even if it is agreed that we cannot access subjective experience from the outside without losing an essential part of this experience (the 'what it is like'-ness) – how do we know we can do justice to the experience while describing it *from the inside out*?

This is where introspection comes in, and it brings us to the second point: introspection is not like 'normal' perception or observation. The subject-object vocabulary that makes sense when we are discussing our perception of things becomes problematic in introspective contexts, where the phenomenon seems to be observing itself. According to a 'direct access' (or 'unmediated observation') model of introspection, this lack of differential distance between subject and object is precisely what endows introspection with an immediate and infallible source of knowledge.[6] Following the footsteps of Descartes (and perhaps Husserl), one might posit the absolute transparency of the self, and infer from this that the most certain knowledge can be derived from introspection.

I believe this model of introspection is the only one that is fundamentally incompatible with my argument below. Since it is widely discredited, I will permit myself to dismiss it on two grounds. First, there appear to be no solid arguments for the notion of introspective self-transparency except perhaps the '*prima facie* evidence' of introspection itself – which is exactly where the problem lies. Second, starting from the premise that there is no subject-object differentiation in introspection as there is in perception, one could argue in either of two ways. One might follow the direct access model in arguing that, wherever subject and object are the same, there is an immediate and therefore certain route to knowledge. But one could just as well take the opposite direction, and argue that where there is no separate object, there is no objectivity, and so the absence of distance is precisely what makes introspection problematic where perception is not.[7] It is this second line of reasoning that I would like to pursue a bit further.

It is possible to argue for any kind of observation that the act of looking changes the object observed. This worry could arise at different levels: in the Kantian consideration that we can never experience the *Ding an sich*, the thing-in-itself that is out of reach of the senses; or in the more trivial notion that we only ever see the world through our own eyes. It rises in a different way in quantum physics: for instance, in Schrödinger's 'cat in the box'- scenario, where it is the very act of looking that determines whether the cat is dead or alive. We could argue, along the transparency-line, that the absence of a basic subject-object distinction makes introspection immune from worries of this kind. Yet we could also argue that the worry becomes more fundamental in this context, as the act of introspection may set up a new and more problematic subject-object distance *within the self*: i.e. the observing-self (subject) versus the self-observed (object). The classic dilemma is whether introspection works through a kind of immediate transparency, or whether it creates a self-as-object that is somehow different from the self-as-subject of our experience.

I do not believe this problem can be solved *a priori* – if it can be solved at all. Fortunately, I do not need to solve it to be able to draw the following conclusion: it is at the very least conceptually and metaphysically *possible* that introspection is not immediate, but is itself a kind of mediation. This possibility is all that is required to present some of the free will philosophers quoted above with the following objection: that they implicitly *and unwarrantedly* assume that their phenomenology and introspection are somehow immediate and therefore constitute a kind of '*prima facie* evidence' to support conclusions about a variety of mental phenomena. I call this the 'phenomenological fallacy': it consists mainly of speaking too easily of 'phenomenal facts'. Considering the possibility of introspective mediation, such implicit trust in active introspection may be dangerous as well as unwarranted. Again, the question must be asked whether the introspected self does justice to the full range of our self-experience.

Here the reader might tug my sleeve, as I appear to be sliding into

dangerous territory. Surely I am not positing an experienced self versus a *real* self – or, to speak in Kantian terms, a phenomenal versus a noumenal self (the self '*an sich*')? On the contrary. I would like to consider the possibility of two levels of 'phenomenal' selves: the self as we encounter it in introspection proper, as opposed to the way we usually experience ourselves. I am interested not in how the self *really* is – after all, there is not much we can say about that – but in how it *usually* is experienced, when we are not actively, philosophically, introspecting.[8] That it is possible for the introspective and experiential self to come apart may not seem intuitively plausible at first. Yet I believe it is possible to realise *through introspection itself* how introspection may distort our self-experience. This may *seem* paradoxical, but it is also unavoidable. That is: we cannot proceed much further here by way of mere argument. Beyond the possibility of mediation mentioned above, we have come to the limits of what we can say about the nature of introspection without invoking some kind of introspection itself. In other words, it is time either to be silent or to introspect. I believe it is worth exploring the second option.

3. The Introspective Argument Against Introspection

Let's take a step backwards. Perhaps we can reconstruct the introspective technique on which the 'common sense' phenomenology of the self and the will is based – the kind that I suspect lies at the heart of the phenomenological fallacy – as follows:

> [INTROSPECTION 1]: Look into yourself. What do you see? You see and feel yourself as an entity of a certain character, with a specific set of desires and beliefs, an irreducible 'I' that is separate from the world. Based on what you want to do and what you think is best to do, you make certain decisions, and realise your intentions into actions. Doing so, you have a 'feeling of doing': you are aware that you have consciously willed these actions, and hence you experience yourself as an agent. Look back at your past actions: usually, when you did something, you intended to do it and then did it. Look forward to your future actions: nothing is stronger than the feeling that it will be *you* who is acting, not merely your body or your brain. Raise your right hand. Did you not feel it was *you* who raised it?

Is this an exaggeration? If it is, it is not meant to be. In the context of this paper, one may already be inclined to take a critical stance to such a leading introspective exercise. But introduced and worded in the right way, such invocations of 'low brow' introspection can have a strong intuitive appeal. To me personally, at least at first sight, the results of [INTROSPECTION 1] seem quite convincing. In the light of actions performed, it seems only natural to think of my agency along the lines of willing-doing: I wanted to go for a walk in the evening, so I decided to do it, and then I did it. Thinking of the future, I can already frame intentions in my mind that will lead to the actions intended. Most strikingly, any version of the hand-example can have the common sense plausibility of Moore's paradigmatic 'Here is one hand'- approach.[9] When I introspect while observing my hand and deciding whether or not to raise it, and then raising it, there does seem to be a 'free-willish' feeling of doing accompanying my movement. Nothing seems more plausible than to say that this is a hand, and this is me raising it.

In themselves, I believe auto-experiments of this kind, in their appeal to common sense experience, are based on a good hunch, a healthy philosophical instinct. But they fail here, since common sense is precisely what is at stake: what *is* our everyday, pre-theoretical experience of ourselves and our actions? What if the kind of self-reflection invoked in [INTROSPECTION 1] is already an interpretation? Here we would do well to remember Wittgenstein's criticism of Moore's commonsensical examples: when philosophers say things like 'I am certain here is a hand' or 'that is a tree', they are already far removed from ordinary language and everyday cases (Wittgenstein, 1969/1975). Similarly, there is no ordinary context in which we would express our conviction that 'I feel it is I who raise my hand', or 'I experience a self', or even 'I have a feeling of doing'. When such things are uttered, we are already deep in theory. Indeed, common sense fails more dramatically here, since it is possible to identify a hand by pointing at it, but it is not possible to simply point at a self or a will.

I think it is relatively easy to discredit [INTROSPECTION 1] as it stands, simply by taking a second look and daring to question its results and assertions. When we are consciously focusing our gaze upon our hand and thinking of raising it, there will indeed be a 'self-centred' feeling of doing. But in normal cases when we are raising a hand, whether to grab something or ask a question or shake another hand, is there a corresponding feeling of deliberate doing – an experience, as Daniel Wegner (2002) would put it, of conscious will? Similarly, in the course of our everyday actions, are we usually actively deliberating and deciding about what to do before doing it? Are we usually *present* in our actions in the way that [INTROSPECTION 1] suggests we are – as consciously wanting, willing, doing selves? For my part, when I reconsider my daily dealings in the world, I think the answer to these questions should be no.

To look at a concrete case, let's consider the action of taking a shower. Wegner uses this example in order to argue that an action cannot qualify as 'truly *willed*' (Wegner, 2002: 3) unless it was accompanied by an experience of conscious will:

> 'If a person plans to take a shower, for example, and says that she intends to do it as she climbs into the water, spends fifteen minutes in there scrubbing up nicely, and then comes out reporting that she indeed seems to have had a shower but does not feel she had consciously willed it – who are we to say that she did will it? Consciously willing an action requires a feeling of doing […], a kind of internal "oomph" that somehow certifies authentically that one has done the action. If she didn't get that feeling about her showering, then there's no way we could establish for sure whether she consciously willed it' (*Ibid.*: 4).

Something appears to have gone wrong here: not just in the notion that it would be natural for us to experience an activity such as showering as consciously willed, and that something important would be missing in the absence of this experience – but in the underlying assumption that it '*usually seems that we consciously will our voluntary actions*', even if '*this is an illusion*' (*Ibid.*: chapter 1, subheading).[10] On the basis of [INTROSPECTION 1], this assumption would indeed seem to be warranted – but is it really?

We could in fact describe the phenomenology of showering in wholly different terms. Speaking for myself, my voluntary showers are not usually preceded by a conscious and deliberate decision, let alone accompanied by an experience of will. Most of the time I don't actively decide to take a shower – I just *do* it. If asked later what I had been doing, I would say 'I took a shower'. If asked why, I might say something along the lines of 'I felt like taking one'. If quizzed by a philosopher, I would indeed explicate that 'I consciously decided to take a shower', and that it was an act of will. But this does not mean that at the moment of showering itself, I had any experience of conscious will at all. On the contrary: as long as I'm not employing a version of [INTROSPECTION 1] in the course of the shower, it's usually a rather passive experience, during which fragments of thoughts and sense input drift in and out of awareness. On the whole, there seems to be no question of a conscious will or decisive self – no 'I' at all – just the experience of showering.

This may also have been what Jean-Paul Sartre had in mind with his notion of the 'transcendence of the ego'. Consider the following example, as worded by Jones & Fogelin (1997):

> 'When I am intensely interested in what I am doing – say, in reading an exciting novel – I never think of myself as reading; I am fully occupied with the narrative. But if, after I have put the book aside, someone asks me what I have been doing, I reply without hesitation, 'I was reading a book.' Where does this knowledge come from? Careful introspection reveals that no 'I' was actually present in my consciousness while I was reading the book. Nevertheless I now know that at that time I was reading. Further, the 'I' that is so seldom present is always available, on call. This too is shown by introspection: I can at any time recall either *what* I experienced on a particular occasion in the past or the fact that it was *I* who experienced it' (371).

Like the shower scenario, this example demonstrates that a different mode of 'careful' introspection, or introspection at *second sight*, can amount to a different kind of self-experience. For lack of a better term, let's call this alternative mode of access to ourselves [INTROSPECTION 2]. Using this mode, it seems to be the case for most voluntary activities – at least, for

most of mine – that we can describe them in different terms from our customary self-reports.[11] In other words, a careful comparative [INTROSPECTION 2] can be wielded to produce a different self-experience than that which results from [INTROSPECTION 1]. This suggests, somewhat paradoxically, that our most commonsensical introspective self-descriptions ('I read a book', 'I wanted to read a book') do not necessarily correspond to our ordinary default self-experience. And that, in a nutshell, is the introspective argument against introspection: philosophers may appeal to introspection in support of a certain notion of self-experience – yet it is introspection itself that can show us that introspection can distort.

Foreground, Background

At this point it may be helpful to introduce the concepts of a foreground and background of experience. I suggest that we may speak of the *foreground* when we are consciously paying attention or thinking about something, and when we are consciously deliberating or deciding. It is here that we may encounter an inner voice, a conscious will, and a centred self. All the rest is in the *background*, which can perhaps be conceived as a dynamic, multi-faceted, often fuzzy web of semiconscious and possibly preconscious footage at the rim of our awareness. Here we may find half-digested thoughts, emotions, intuitions, perceptions, and even actions: things that are not interpreted by the foreground, but are nevertheless part of our experience, and can be stored for future use.[12]

It should be noted that this schematic distinction between a foreground and a background is *not* meant to correspond to the dichotomy of conscious versus unconscious or automatic behaviour, which is often presupposed by philosophers, psychologists and scientists alike. For instance, psychologists such as Bargh & Chartrand (1999) discuss an array of empirical evidence to argue that 'most of our day-to-day actions, motivations, judgments, and emotions are not the products of conscious choice and guidance', but are 'driven by automatic, nonconscious mental processes' – indeed, 'it appears impossible, from these findings, that conscious control could be up to the job' (464-5). Underlying their thesis is a clear-cut distinction between conscious processes on the one hand, which are associated with awareness, intent, effort and control, and nonconscious processes or automatisms on the other (463). To me, such a black-and-white distinction of conscious control versus nonconscious automatism seems deeply flawed, as it overlooks the wide grey zone of pre- and semiconscious aspects of experience.[13] Consequently, it makes the primacy of the nonconscious seem counter-intuitive, as it stipulates that such processes pass outside the reach of our experience, awareness and agency.

This does not have to be the case. What the concept of a background may help us to grasp intuitively, is that even if an action is not actively, consciously *willed*, this does not mean it is not part of our awareness, our experience – of *who we are*. The boundaries between foreground and background should be visualised as always flowing, shifting, oscillating, fluctuating: whether continuously, as in James' concept of a 'river' or 'stream' of consciousness (1890/1950: 239), or discontinuously, 'constantly broken by detours – by blows – fissures – white noise' (Strawson, 2003: 356). The difference between the grounds is gradual, dynamic, and unstable: what is now in the foreground may merge into the background before we know it, and bits and pieces of the background may pop up into the foreground and evaporate again in the blink of an eye.[14] Actions may flow from the background as well as the foreground, and this does not disqualify either kind from being rightly attributed to our (possibly retrospective) sense of agency.

Hence, I do not deny that the foreground is a real and important part of our experience. Indeed, it could be argued that our most essential and self-defining actions spring from the foreground – for instance, when we are facing difficult decisions or moral dilemmas, when we are trying to figure out what course of action would be best, or when we are consciously reflecting on our behaviour. But even if the foreground is a real and important *part* of our lives, this does not mean it is the *only* part. I believe experience can show us that most of the time, we are living in the background, although the foreground is always on call. In fact, the very moment we want to take a closer, more conscious look at things, the shift is made, and we are already in the foreground. And this is why we cannot trust the kind of active introspection some philosophers endorse. For the problem with any variety of [INTROSPECTION 1] is that it is a child of the

foreground, and speaks no other language than that of a conscious, willing self. Hence, whenever we introspect or 'retrospect', any experience that usually belongs to the background is drawn into the foreground, and interpreted in terms of a one-dimensional self. Of course nothing is wrong with this in principle: let all introspect as they please. It is not until philosophers start drawing conclusions that things get messy.

My thesis is that [INTROSPECTION 1] by nature only interprets the foreground, *even where there was none* – and therefore, that we need an alternative along the lines of [INTROSPECTION 2] to get in touch with the background, which makes up a more integral part of our everyday experience. Here we appear to meet a paradox. For if, as I have suggested, the experience of the self is mediated by the reflective act, how can we reflect on our usual experience without such mediation? How can we use introspection to get behind introspection? Perhaps we should accept this paradox, since we cannot unravel the Kantian knot: we cannot access our experience except through our experience, and we cannot see the self without looking at it. But my claims are more modest than that. My point is rather that different ways of introspecting lead to different (and sometimes incompatible) introspective results. Furthermore, I believe the kind of mediation that comes with [INTROSPECTION 1] can be avoided to some extent by employing the 'method of stealth' of [INTROSPECTION 2], as I have tried to do in the shower example.

This consists less in active introspection than in a more passive procedure of monitoring our daily experience, to catch us 'in the act' of consciousness. One could say that instead of switching on the light and shouting 'freeze!', we could try to sneak up on ourselves in the dark, and thus hope to catch a glimpse of ourselves before the limelight of introspection is on. We cannot completely avoid the focused, deliberate introspective gaze, for the light must go on at some point. When it does, the phenomenon as it was (an unreflected embeddedness in the background) disappears, as the background is irrevocably drawn into the focus of the foreground and loses its fuzzy, unfocused character. But traces may remain, may be remembered. If nothing else, [INTROSPECTION 2] may prove useful in moderating the results produced by [INTROSPECTION 1], and supplementing them with its own.[15]

Conclusion

Summing up, there is a major risk in trusting any kind of introspection at face value, and there lies a phenomenological fallacy in relying too easily upon 'phenomenal facts'. At the same time, phenomenology cannot go without the first-person perspective: any attempt to formulate a third-person phenomenology retains an element of naivety, since it fails to recognise that third-person descriptions ultimately go back to a first-person stance. If phenomenology is to survive at all, some kind of introspection is required. The question is, of course, *what* kind – since the chosen 'method' of introspection may determine the results. In the context of this paper, [INTROSPECTION 2] was used to demonstrate the possible deficits of [INTROSPECTION 1]. The latter often boasts of possessing an unproblematic, commonsensical, *prima facie* lucidity, yet it remains to be seen on a case-by-case basis whether it has enough explanatory power to uphold its bold conclusions. This is why it is healthy at times to use a version of [INTROSPECTION 2], to bring us as close to usual experience as possible, and at least try to look at it from the background's point of view.

What, then, does this imply for free will philosophers? In my view, several do's and don'ts. First of all, do introspect. And second, do draw the background into it. But third, don't universalise your introspection; and fourth, don't automatically equate it with your usual experience. Fifth and finally, don't draw too clear a line between the conscious and the nonconscious – for there may be a whole realm of experience in between. Perhaps, if these lessons are learned, we can question the persistent *reductive* conception of the will and the self, among proponents of mental causation and their critics alike, as something that remains once we have subtracted from our behaviour all nonconscious (environ)mental factors that have any causal influence upon it. The main question is often considered to be whether such a conception makes metaphysical sense, when the real question should be whether it makes *experiential* sense.

I do not believe this is how we usually experience our agency, nor our identity. We do not just identify with the beam of active, deliberating, focused consciousness that is most often projected onto the philosophical stage. We also identify with the 'semi-automatic' stream of thoughts, actions and impressions, which are not (fully) conscious to us, but are

nevertheless *there*: the many things in the background that give meaning and shape to the foreground, and deserve not to be overlooked. If the self consisted only in the truly conscious part, not much would remain of it: it would not contain, for instance, the sudden flashes of insight, inspiration, or creativity, which appear to rise from nowhere, yet are so essentially part of who we are. Nor the kind and cruel acts we do without thinking, and afterwards may cherish, or regret. Such things may not be part of the focal point of our experience, but they are part of its horizon.

Trying to catch these fragile fragments of experience may take us to the limits of language, since we will have to resist the temptation to speak of the background in terms of the foreground. As Searle noted (1983: 157): 'The price we pay for deliberately going against ordinary language is metaphor, oxymoron, and outright neologism.' But what we lose in clarity, we gain in lifelikeness: for if the talk of backgrounds seems fuzzy, whoever said that experience was not?

Mara van der Lugt (1986) studied Philosophy at Erasmus University Rotterdam, and obtained her master's degree cum laude in 2010. For her second master in Early Modern Intellectual History she received the Prof. Bruins Prize for best research master student at Erasmus University in the year 2010. She is currently pursuing a PhD in History of Philosophy on the writings of the 17th-century French philosopher Pierre Bayle, at Corpus Christi College, University of Oxford.

'The Foreground and Background of Consciousness' was written for the mastercourse 'Experimenting Ethics Away' (of dr. Maureen Sie), taught by dr. Leon de Bruin.

Literature

Bargh, J.A., & Chartrand, T.L. (1999) 'The Unbearable Automaticity of Being'. In: *American Psychologist* 54, no. 7, 462-479.

Bayne, T. (2006) 'Phenomenology and the Feeling of Doing: Wegner on the Conscious Will'. In S. Pockett, W.P. Banks & S. Gallagher (eds.) *Does Consciousness Cause Behavior?* Cambridge, MA: MIT-Press, 169-186.

Bayne, T. (2004) 'Self-consciousness and the Unity of Consciousness'. In: *The Monist* 87, no. 2, 219-236.

Bradley, R. D. (1958) 'Free Will: Problem of Pseudo-Problem?'. In: *Australian Journal of Philosophy* 36, no. 1, 33-45.

Brown, S.R. (2000) 'Tip-of-the-Tongue Phenomena: An Introductory Phenomenological Analysis'. In: *Consciousness and Cognition* 9, 516-537.

Dennett, D.C. (1991) *Consciousness Explained*. London: Penguin Books.

Dennett, D.C. (1984) *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, MA: MIT Press.

Gertler, B. (2008) 'Self-Knowledge'. In E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2008 Edition). Available from: plato.stanford.edu/archives/win2008/entries/self-knowledge [30 November 2011].

Ginet, C. (1990) *On Action*. Cambridge: Cambridge University Press.

Gurwitsch, A. (1964) *The Field of Consciousness*. Pittsburgh: Duquesne University Press.

Hohwy, J. (2004) 'The Experience of Mental Causation' In: *Behavior and Philosophy* 32, 377-400.

James, W. (1890/1950) *The principles of psychology*. New York: Dover Publications.

Jones, W.T. & Fogelin, R.J. (1997) *The Twentieth Century to Quine and Derrida*. Fort Worth: Harbourt Brace College Publishers.

Kane, R. (2005) *A Contemporary Introduction to Free Will*. New York/Oxford: Oxford University Press.

Lewis, M. & Staehler, T. (2010) *Phenomenology: an introduction*. London/New York: Continuum International Publishing Group.

Libet, B. (1999), 'Do We Have Free Will?' In: *Journal of Consciousness Studies* 6, no. 8-9, 47-57.

Moore, G.E. (1962) *Philosophical Papers*. New York: Collier Books, 1962.

Moran, R. (2001) *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton/Oxford: Princeton University Press.

Nahmias, E. (2002) 'When consciousness matters: a critical review of Daniel Wegner's The illusion of conscious will'. In: *Philosophical Psychology* 15, no. 4, 2002.

Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2004) 'The Phenomenology of Free Will'. In: *Journal of Consciousness Studies* 11, no. 7-8, 162-79.

O'Connor, T. (1995) 'Agent Causation'. In T. O'Connor (ed.) *Agents, Causes, and Events*. New York: Oxford University Press.

Polanyi , M. (1983) *The Tacit Dimension*. Gloucester, Mass.: Peter Smith.

Searle, J.R. (1983) *Intentionality, an essay in the philosophy of mind*. Cambridge: Cambridge University Press.

Searle, J.R. (1984) *Minds, brains, and science*. Cambridge, MA: Harvard University Press.

Smith, D.W. (2011) 'Phenomenology'. In E.N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2011 Edition). Available from: plato.stanford.edu/archives/fall2011/entries/phenomenology [30 November 2011].

Strawson, G. (2003) 'The Self'. In R. Martin & J. Barresi (eds.) *Personal Identity*. Oxford: Blackwell Publishing Ltd. 2003.

Wegner, D.M. (2002) *The illusion of conscious will*. Cambridge, MA: MIT Press.

Wittgenstein, L. (1951/ 2007) *Philosophical Investigations*. ed. G.E.M. Anscombe. Oxford: Blackwell Publishing Ltd.

Wittgenstein, L. (1969/1975) *On Certainty*. ed. G.E.M. Ansombe/ G.H. von Wright, Oxford: Blackwell Publishing Ltd.

Whitehead, A.N. (1947) *Essays in Science and Philosophy*. New York: Philosophical Library.

Notes

1 Quoted in Bradley (1958: 38).

2 Phenomenology can be defined as 'the study of structures of consciousness as experienced from the first-person point of view.' (Smith, 2008: unpaginated). 'Phenomenology does not attempt to speak about things, but only about the way they manifest themselves, and hence it tries to describe the nature of *appearance as such*.' (Lewis & Staehler, 2010: 1). In the context of this paper 'phenomenological' designates any attempt to describe consciousness and experience from the first-person perspective. (See section 2 below.)

3 E.g. claims about experience that pretend to apply to all humans, such as: 'Everyone experiences the world as unified', or: 'We all have an experience of conscious will.'

4 The free will debate may also be construed as a conflict between determinism and the demands of moral responsibility, in which case experiential factors can be (but are not always) left out. For a general overview of the free will debate, see for instance Kane (2005).

5 Compatibilists (who believe free will to be compatible with determinism) appear to be less committed to a 'self as source'- phenomenology, and even tend to 'describe the deliberative process more passively, with our decisions "flowing from" our desires and beliefs' (Nahmias *et al.*, 2004: 167) – yet they do sometimes imply the kind of mental causation phenomenology that is described by Hohwy above. Though I will here mainly question the self-phenomenology implied by libertarians (who believe determinism to be false) and their opponents, this is not to say compatibilists are not at all implicated in my criticism below, as it is directed at any kind of unquestioned introspection.

6 A version of this, as phrased by Gertler (2008: unpaginated): 'Since an appearance of a phenomenal quality and the reality which appears (the phenomenal quality itself) are one and the same, on this account, one can enjoy epistemically direct access to the phenomenal quality by attending to it.'

7 See for instance Moran (2001: 27-8): '[...] the problem of self-knowledge is not set by the fact that first-person reports are especially good or reliable, but primarily by the fact that they involve a distinctive mode of awareness, and that self-consciousness has specific *consequences* for the object of awareness.'

8 Note that in this context 'usually' does not necessarily imply 'most of the time': rather the 'usual' experience of the self signifies a default mode of consciousness prior to introspection. Though in practice it is possible we are in this default mode most of the time, this is not *necessarily* so.

9 Moore famously delivered a 'Proof of an External World' by means of raising one hand, then another, in order to show how an appeal to common sense can solve or dissolve philosophical dilemma's (Moore, 1962: 144-8). Hand-raising examples appear here and there in the free will debate, as in Bayne (2006: 176) and, interestingly, in the experimental context of Libet 1999, where 'the sudden flick of the wrist' is considered a typical act of will (50). See also Wittgenstein (1953/ 2001: no. 621): 'Let us not forget this: when "I raise my

arm", my arm goes up. And the problem arises: what is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?'

10 Note that we can detect two kinds of universalising tendencies in Wegner's argument: not just the universalisation from 'I' to 'we' as discussed above, but also a universalisation from one or a few instances of introspection to a wide range of human experience: from 'now' to 'most of the time'. Searle seems to do the same in claiming that we experience free will 'all the time' (1984: 88). Though my argument is directed against any failure to discriminate between experience *pre-* and *post*-introspection, this second kind of unwarranted universalisation gives it particular urgency.

11 Consider Dennett's alternative phenomenology of decision-making: '[...] decisions can also be seen to be strangely out of our control. We have to wait to see how we are going to decide something, and when we do decide, our decision bubbles up to consciousness from we know not where. We do not witness it being *made*: we witness its *arrival*.' (1984: 78).

12 This notion of a background to experience is not the same as the background to belief as conceptualised by Searle (1983), or the 'tacit dimension' of Polanyi (1983), though they may be related. Gurwitsch comes close, yet is too static in his division of the 'field of consciousness' into the '*theme*' or focus of attention, the '*thematic field*' or relevant background to the theme, and the irrelevant data in the '*margin*' (1964: 4). More flexible is James' concept of a '*psychic overtone, suffusion,* or *fringe*, to designate the influence of a faint brain-process upon our thought, as it makes it aware of relations and objects but dimly perceived' (1950: 258). Whitehead also speaks of 'the dim background of our conscious experience' (1947: 122).

13 These include things of which you are not really conscious at the moment itself, though in retrospect you may realise they were in your awareness. For instance: a neighbour is playing some nice classical music. It is not until the music stops that you realise you were listening and enjoying it the whole time. It may have been in the background – but it was there. One might also think of actions that were not accompanied by a conscious intentional stance, though in retrospect one would appropriately ascribe an intention to them.

14 Tip-of-the-tongue phenomena are interesting in this context, as they may make us aware of an absence or background to experience (Brown, 2002); of 'a gap that is intensely active' (James, 1890/1950: 251).

15 There still seems to be a paradox in the fact that I too am drawing on my own introspective experience as a basis for my argument, and I can prove neither that I am correctly describing my experience, nor that this description would match the experience of others. Again, I accept this paradox, firstly because I believe it to be unavoidable in any phenomenological discussion, and secondly because I do not mean to propose an alternative universalistic account of the experience of conscious will. I am trying not to generalise, but to *de-generalise*: to question the results of one common kind of introspection by exposing them to a second, more comparative kind. Perhaps my use of introspection should likewise be questioned and reconsidered: in fact my argument would encourage precisely such a re-examination.

ESJP
#1 | 2011

Het spel der betekenissen – Een speelse hermeneutiek van de hermeneutiek

Elize de Mul

La folie est de penser trop de choses dans une succession trop rapide, ou d'une chose trop exclusivement.

– Voltaire

Ongeveer één keer in de week, meestal op een midweekse dag, word ik ruw in mijn dagelijkse bezigheden gestoord door het rinkelen van de deurbel. Gezien het vreemde tijdstip – rond een uur of drie in de middag – waarschuwt een stemmetje in mijn achterhoofd de deur met rust te laten. Tevergeefs. Eenmaal in de deuropening blijkt mijn voorgevoel meestal juist en wordt er onmiddellijk met groot enthousiasme een pamflet van deze of gene kerkgemeenschap in mijn handen gedrukt (door een speling van het lot bevindt mijn appartement zich in de nabijheid van maar liefst acht kerkgemeenschappen, waarvan het merendeel zeer actief is op het gebied van zieltjeswinning). Ondanks de – volgens de pamfletten – vaak zeer uiteenlopende geloofsovertuigingen, claimen al deze dames en heren te verkondigen 'wat er letterlijk in de Bijbel geschreven staat'. Een gesprek aangaan met deze ongenode gasten blijkt keer op keer een moeilijke, zelfs onmogelijke aangelegenheid. Welke goedgefundeerde argumenten ik ook inbreng in de hoop een discussie te ontlokken, tegen de logica van de Heer blijken zij niet bestand. Een gelovige monoloog is het gevolg, waarbij alle inbreng van mijn kant geheel vanuit het licht van de Bijbel wordt geïnterpreteerd. Frustrerend, althans voor mij. De 'interpretatiekloof' die ik bij elk van deze aanvaringen ervaar, is te herleiden naar het moment in de geschiedenis van de Westerse cultuur waarop God sommigen verliet en hen zonder spiegel Gods in een staat van verwarring achterliet, met enkel hun naakte zelf als referentiekader. Inmiddels zijn die 'sommigen' in een land als Nederland een meerderheid en gezien mijn in het bovenstaande beschreven frustratie mag het duidelijk zijn dat ik mij onder hen mag scharen. Gelukkig voor ons waren daar al snel enkele, geestelijk goedbedeelde, individuen (want dat waren we ineens) die ons wisten te kleden met filosofische overdenkingen en ons een nieuwe plek in de wereld en ten opzichte van onszelf wisten aan te meten. Daar God niet meer het vanzelfsprekende kader is waar vanuit de wereld begrepen dient te worden, spelen er ineens allerlei, vaak verwarrende vragen over de wijze waarop wij de wereld dán wel kunnen verstaan en over wat dit hele 'verstaan' nou eigenlijk is.

Al vanaf het moment dat je (al dan niet bewust) waarneemt, ben je in feite aan één stuk door bezig om op meer of minder geslaagde wijze de wereld om je heen te interpreteren. Interpreteren is een zeer belangrijke bezigheid. Niet alleen voor de interpreet, die hiermee de chaotische en gefragmenteerde werkelijkheid leefbaar maakt, maar evengoed ook voor het geïnterpreteerde. De snedige en sarcastische opmerking van je baas staat of valt met jouw begrip van zijn toon. Een verkeersbord oefent slechts invloed uit op je rijstijl wanneer je het correct 'begrijpt' (eventuele ongelukken door een al te langdurig staren daargelaten). Deze voorbeelden lijken te impliceren dat er per geval een 'juiste' interpretatie bestaat. Daarbij spelen onder andere geschreven en ongeschreven sociale regels (waarover de stilzwijgende afspraak bestaat dat we ons daar aan dienen te houden), conventies en ervaring een grote rol. In het dagelijks leven wordt hier echter niet erg vaak bij stil gestaan. Meestal gebeurt dit pas in gevallen van 'verstaansverwarring', bijvoorbeeld wanneer iemand de (spel)regels niet lijkt te volgen. Behalve in het alledaagse leven speelt de praktijk van het verstaan (hermeneuse) ook in de wetenschap een belangrijke rol. Interpretatie is de belangrijkste methode binnen de geesteswetenschappen en neemt ook in

de kwalitatieve sociale wetenschappen en zelfs in de natuurwetenschappen een belangrijke plaats in. De hermeneutiek heeft zich van oudsher ontwikkeld als een *methodologie*, gericht op het opstellen van regels voor de interpretatie. De filosofische hermeneutiek, ten slotte, richt zich op een interpretatie van het verstaan zélf en reflecteert daarbij op het gegeven dat de mens een interpreterend wezen is.

Niet alleen het alledaagse 'verstaan' wordt gekenmerkt door geschreven en ongeschreven (spel)regels, datzelfde geldt ook voor de *filosofische* interpretatie van het 'verstaan'. Verschillende tradities binnen de filosofische hermeneutiek veronderstellen verschillende 'regels' van 'verstaan'. In het volgende wil ik laten zien dat de speltheorie, zoals die onder anderen aan de hand van denkers als Huizinga en Callois tot ontwikkeling is gekomen in de *New Media Studies*, zowel tot verheldering als verdieping van de verschillende tradities binnen de hermeneutiek kan dienen. Mijn betoog zal zich dus voornamelijk afspelen op het vlak van de filosofische hermeneutiek. Hierbij begeef ik me op een metaniveau en presenteer ik dus als het ware een hermeneutiek van de filosofische hermeneutiek.

Het lijkt wellicht wat wonderlijk een op het eerste gezicht 'ernstige' aangelegenheid als het interpreteren van de werkelijkheid te verbinden met een 'frivool' verschijnsel als spel. Hermeneutiek en spel, zo zal ik betogen, blijken echter meer met elkaar van doen te hebben dan men op het eerste gezicht zou denken. In het volgende zal ik bekijken wat voor een rol het 'spel' inneemt in het 'verstaan'. Hiertoe bespreek ik eerst – kort – de geschiedenis van de hermeneutiek waarbij ik de door Jos de Mul geschetste driedeling - reconstructie, constructie en deconstructie - zal hanteren (De Mul, 1993). Na deze korte inleiding in de hermeneutiek volgt een analyse van de rol die het 'spel' speelt in de hermeneutiek, waarbij ik gebruik zal maken van de speltheorieën van Huizinga, Caillois en Gadamer. Deze analyse voltrekt zich voor een klein deel in de praktijk van het 'verstaan' zelf (de hermeneuse) maar vooral ook op het theoretische vlak van de hermeneutiek. Ten slotte introduceer ik Derrida als 'spelbreker' in mijn verhaal en betoog ik waarom 'spelbreken'– anders dan Huizinga van mening is – juist tot vernieuwing van het verstaan kan leiden. *Let the game begin*.

Van reconstructie tot deconstructie

De geschiedenis van de hermeneutiek neemt in de Westerse filosofie een aanvang wanneer er een bewustzijn ontstaat van de historische en culturele begrensdheid van het menselijk kenvermogen. De filosofie van Immanuel Kant (1724-1804) speelde hierbij een belangrijke rol. Weliswaar krijgt het historisch en cultureel bepaalde individu in zijn werk nog weinig aandacht – in zijn filosofie gaat het namelijk om *de* mens, die wordt opgevat als een tijdloos, abstract subject, dat wordt gekenmerkt door 'een uniforme en universele menselijke natuur' (De Mul, 1993: 131) –, toch lijkt ergens al het besef van de menselijke eindigheid bij hem te broeden; wat tijdens zijn kritische periode (vanaf 1781) steeds explicieter tot uitdrukking komt. Hiermee zet hij de toon voor latere filosofische bespiegelingen over de mens. De opvatting dat de mens en de maatschappij product zouden zijn van een uniforme en universele menselijke natuur verliest dan ook aanhang in de post-kantiaanse filosofie, onder andere door de 'ontdekking' van niet-westerse culturen (De Mul, 1993: 132). Bovendien wordt in de loop van de negentiende eeuw, als gevolg van een toenemende 'historisering van het wereldbeeld', de notie van tijdloosheid geproblematiseerd. Deze radicalisering van het historisch besef 'leidt echter onafwendbaar tot een fundamentele betwijfeling van de onaantastbare status van de tegenwoordige rationaliteit en daarmee tot een historisch relativisme, dat tot op heden de historische reflectie doortrekt' (De Mul, 1993: 132). Het historisch bewustzijn is daarom in zekere zin tragisch te noemen, daar het wetenschappelijk verlangen naar definitieve interpretaties door dit besef wordt gesaboteerd (De Mul, 1993: 209).

Enkele filosofische bewegingen uit deze tijd van plotselinge vertwijfeling zijn het idealisme en het positivisme. Ondanks het feit dat ze in veel opzichten radicaal tegenover elkaar staan, worden beiden gekenmerkt door een sterk vooruitgangsgeloof. Het is voornamelijk dit laatste waartegen Duitse historisten zich af gaan zetten, met de romantische traditie als grote inspirator (De Mul, 1993: 152). De romantische hermeneutiek van Friedrich Schleiermacher (1768-1834) is hiervan een goed voorbeeld. Schleiermacher blijft, zoals veel romantici, trouw aan de 'eindige mens' van de latere Kant. Hij lijkt deze zelfs te radicaliseren, wanneer hij Kant verwijt 'de menselijke rede op een a-historische wijze te benaderen en de rol van

taal te veronachtzamen' (De Mul, 1993: 153). Als gevolg hiervan treedt individualiteit in Schleiermacher's hermeneutiek op de voorgrond en definieert hij het verstaan als een gesprek tussen individuen. Maar hoewel het verstaan door Schleiermacher wordt gepresenteerd als een dialogische relatie tussen individuen, kan gesteld worden dat de door hem voorgestane hermeneutiek feitelijk eerder een 'monoloog' is. De gesprekspartner 'fungeert' in zijn theorie toch vooral als 'object', waarvan de betekenis moet worden *gereconstrueerd*. Dit komt duidelijk naar voren in zijn opvatting dat men een auteur beter zou kunnen verstaan dan hij zichzelf verstaat. Dit lijkt tegenstrijdig te zijn met zijn notie van individualiteit en de gedachte dat we nooit volledig tot de individualiteit van een ander kunnen doordringen (De Mul, 1993: 162). Echter, Schleiermachers idee dat we een auteur beter zouden kunnen verstaan dan hij zichzelf verstaat, sluit aan bij het concept van de 'hermeneutische cirkel van het verstaan'. Schleiermacher stelt dat een woord geen vaste betekenis heeft, maar dat deze betekenis afhankelijk is van de contexten waarin het wordt gebruikt. Wanneer de context wordt uitgebreid, krijgt een woord vaak een bredere betekenis. Hiermee is de hermeneutische cirkel niet zozeer een gesloten kring, maar eerder een opwaartse spiraal, waarbij de betekenis van te interpreteren teksten met de uitbreiding van de context steeds verder wordt verrijkt met nieuwe inzichten (De Mul, 1993: 162). Vandaar dat het mogelijk is een auteur 'beter' te verstaan dan hij zichzelf verstaat, daar de betekeniscontext alsmaar rijker wordt. Toch blijft interpretatie bij Schleiermacher letterlijk eenzijdig, daar de betekenis van een 'tekst' geheel afhangt van de interpretatiecontext van de interpreet.

Ook Wilhelm Dilthey (1833-1911) blijft Kant in veel opzichten trouw, onder andere door in zijn 'bewustzijnsfilosofische voetsporen' (De Mul, 1993: 172) te treden. Dit houdt in dat kennis zowel bij Kant als bij Dilthey betrekking heeft op de voorstellingen van een 'kentheoretisch subject' (De Mul, 1993: 172). Dilthey voegt hier echter wel een belangrijke kritische noot aan toe, door Kant te verwijten dat hij zich, wanneer hij over ervaring spreekt, beperkt tot de ervaring van de *fysische* werkelijkheid. Dilthey stelt dat de menselijke ervaring zich niet enkel beperkt tot natuurlijke verschijnselen (*Erscheinungen*), maar ook deels stoelt op geestelijke verschijnselen (De Mul, 1993: 173). Daarmee doelt hij niet alleen op psychische inhouden, maar vooral ook op de materiële producten van de geest (handelingen, boeken, gebouwen etc.). Deze materiële producten met een geestelijke lading vormen samen de historisch-maatschappelijke werkelijkheid. Belangrijk hierbij is dat materie en geest hier niet streng te scheiden zijn. Sterker, Dilthey stelt dat natuur- en geesteswetenschappen 'niet primair betrekking [hebben] op verschillende delen van de werkelijkheid, maar zijn gebaseerd op verschillende ervaringswijzen van een zelfde werkelijkheid' (De Mul, 1993: 176). Natuurwetenschappen richten zich uitsluitend op de wereld zoals die in de uiterlijke ervaring (d.w.z. door de zintuigen) is gegeven, en proberen hetgeen ervaren wordt door middel van wetmatigheden te *verklaren*. De geesteswetenschappen kenmerken zich door een proces, dat Dilthey in navolging van Schleiermacher aanduidt als *verstaan*. Dit verstaan is een proces waarbij de uiterlijke ervaring met de innerlijke ervaring wordt verbonden. Dilthey verbreedt Schleiermacher's notie van verstaan doordat hij zich niet beperkt tot het verstaan van talige uitdrukkingen, maar de gehele historisch-maatschappelijke werkelijkheid tot object voor het verstaan bestempelt. Dit betekent dat het kentheoretisch subject een bijzondere dubbelrol speelt, omdat het zowel subject als object van de geesteswetenschappen is. Dilthey onderstreept dat er niet zoiets bestaat als een 'onschuldig verstaan'. De Mul:

> 'As we always start from our own finite horizon, we are always tempted to solely interpret the other culture in terms of our own finite horizon and in doing so to reduce the 'other' to the 'same'. Moreover, in these cases the own culture is often conceived of as being superior to the foreign culture' (De Mul, 2011: 11).

Verstaan gaat hierdoor vaak gepaard met 'geweld', waarbij een 'tekst' een bepaalde interpretatie 'opgedrukt' krijgt. Doordat het proces van verstaan wordt gekenmerkt door een verbintenis van uiterlijke *en* innerlijke wereld, is de historisch-maatschappelijke wereld verder niet eenduidig te verklaren en te begrijpen. Dilthey radicaliseert, onder andere door deze gedachtegang, de door Kant als eerste gethematiseerde eindigheid van de mens en is hiermee een belangrijke inspirator voor latere hermeneutici als Martin Heidegger (1889-1976) en Hans-Georg Gadamer (1900-2002) (De Mul, 2011: 11). Gadamer is in grote mate schatplichtig aan Heidegger, maar aangezien hij in zijn filosofische hermeneutiek, anders dan

Heidegger, het thema spel een belangrijke plaats geeft, beperk ik mij hier tot zijn werk.

Eerder werd al gesteld dat hoewel Schleiermacher het proces van verstaan voorstelt als een gesprek, dit in de praktijk eerder een monoloog bleek te zijn. Dat geldt in belangrijke mate ook nog voor Dilthey, wat onder andere tot uitdrukking komt in zijn metafoor van de 'horizonsverbreding' (De Mul, 2011: 3). De metafoor van een gesprek blijkt eerder van toepassing te zijn op de hermeneutiek van Gadamer. Interpretatie kan volgens Gadamer worden gezien als een 'horizonsversmelting', ofwel een *constructie* van betekenis (al stelt hij zelf al dat deze metafoor, die impliceert dat er twee duidelijk afgebakende horizonnen zijn, al te simplistisch is (De Mul, 1993: 13)). Interpreteren is niet zozeer gericht op theoretische kennis, maar is eerder een praktische aangelegenheid. Het doel is niet om een oorspronkelijke betekenis te reconstrueren, zoals bij de reconstructieve hermeneutiek het geval is, maar om een tekst gericht te benaderen met vragen uit het heden om zodoende een vruchtbare 'dialoog' aan te gaan die wellicht tot een relevant inzicht kan leiden. Wanneer dit gebeurt is er sprake van een horizonsversmelting: zowel de horizon van de interpreet als die van het geïnterpreteerde zijn niet meer geheel hetzelfde als vóór de interpretatie. Zo kan een gebouw ons bijvoorbeeld veel leren over een bepaalde cultuur, terwijl deze waarschijnlijk niet per se met dit doel is ontworpen en gebouwd (De Mul, 1993: 331). Hoewel Gadamer zich met zijn notie 'horizonsversmelting' tegen het interpreteergeweld van de 'horizonsverbreding' van Schleiermacher en Dilthey keert, is alleen al in dit voorbeeld te zien dat ook dit 'gesprek' niet altijd evenwichtig is. Het gevaar hiervan zou kunnen zijn – zoals onder andere Derrida opmerkt, en zoals we later zullen zien – dat ook een dergelijke manier van interpreteren in de praktijk met veel interpretatief geweld gepaard gaat. De interpreet heeft de neiging om zijn eigen visie al te veel aan het geïnterpreteerde op te leggen waardoor dit teveel, of zelfs zijn hele, oorspronkelijke betekenis verliest. Concluderend blijkt bij zowel constructieve- als reconstructieve hermeneutiek het gevaar van een al te eenzijdige interpretatie te blijven bestaan.

Of Jacques Derrida (1930-2004) tot de hermeneutiek gerekend kan worden, wordt wel betwijfeld. Hij duidt zijn werk niet met die term aan en sommigen zijn zelfs geneigd zijn 'deconstructieve praktijk' als een 'anti-hermeneutiek' te bestempelen (De Mul, 2011: 11). Anderzijds is duidelijk dat zijn werk veel gemeen heeft met de hermeneutische traditie. In feite radicaliseert Derrida de theorie van Schleiermacher, die stelt dat een woord geen gefixeerde betekenis heeft, maar dat deze betekenis afhangt van de context waarin het woord wordt geïnterpreteerd en daarmee steeds kan veranderen. Een dergelijke 'ervaringshorizon' kan met andere woorden in principe naar alle kanten oneindig worden uitgebreid. Derrida concludeert hieruit dat interpreteren een 'oneindige taak' is door te stellen dat bij elke interpretatie van een 'tekst' er een beslissing wordt genomen over iets dat in feite fundamenteel 'onbeslisbaar' is (De Mul, 2011: 11). Verder kunnen woorden volgens Derrida uit hun context worden gehaald en in een nieuwe context worden geplaatst, waardoor er naast de notie van de 'onbeslisbaarheid' van een interpretatie ook nog eens een oneindig aantal nieuwe betekenissen kan worden gegenereerd. Om bij de metafoor van een 'verstaanshorizon' te blijven, kunnen we hierdoor stellen dat er bij Derrida sprake lijkt te zijn van een extreme 'horizons*verstrooiing*'. Op deze deconstructieve hermeneutiek van Derrida zal ik verderop nog nader ingaan.

Spelen met taal

Om poëzie te verstaan,
moet men de ziel van het kind kunnen aantrekken als een tooverhemd
en de wijsheid van het kind aanvaarden boven die van den man.

– Johan Huizinga

De eerste keer dat Johan Huizinga (1872-1945) wat langer stilstaat bij het thema 'spel' is in zijn rectorale rede 'Over de grenzen van spel en ernst in de cultuur' uit 1933. Hij noemt hier het concept spel reeds een 'categorie die alles verslindt' (geciteerd door Otterspeer in Huizinga, 2008: z.p.). Dat het spel inderdaad alles, inclusief zijn aandacht verslindt, blijkt ook uit feit dat het onderwerp hem niet meer los laat en hij vijf jaar later zijn bekende boek *Homo Ludens: Proeve eener bepaling van het spel-element der cultuur* publiceert. Hij betoogt hierin dat spel niet zozeer *onderdeel* is van

de menselijke cultuur, maar dat cultuur eerder *uit het spel* lijkt te ontspringen. Door het spel als grondveste voor de gehele cultuur op te vatten wordt aan het kentheoretisch subject dat vanuit een bepaalde context het concept tracht te begrijpen een zelfde soort dubbelrol toegekend die wij eerder al waarnamen bij Dilthey. Huizinga:

> 'Wie het oog richt op de functie van het spel […] vindt het spel in de cultuur als een gegeven grootheid, bestaande vóór de cultuur zelve, haar begeleidend en doortrekkend van begin af tot in de phase van cultuur, die hij zelf beleeft' (Huizinga, 2008: 31).

Hoewel het niet zijn doel is een eenduidige definitie te geven van het begrip 'spel' of een verklaring voor de vraag 'waarom men speelt', vangt Huizinga in *Homo Ludens* wel aan met een poging te vatten wat de voorwaarden zijn voor spel. Hij noemt een achttal kenmerken van het spel op. Zo is het spel bijvoorbeeld op te vatten als een vrije handeling, die zich niet gemeend en buiten het gewone leven staande afspeelt binnen een bepaalde tijd en ruimte en die volgens bepaalde regels verloopt (Huizinga, 2008: 41). Waar het zwaartepunt van Huizinga's verdere betoog ligt op zijn analyse van hoe cultuur *in* en *als* spel ontstaat, is het vreemd genoeg voornamelijk deze – volgens hemzelf terloopse – poging het spel te definiëren die door andere denkers veelvuldig is aangewend om (onder andere) *games* mee te analyseren. Ik zeg hier 'vreemd', omdat de ondertoon van *Homo Ludens* op zijn minst kritisch te noemen is, en er aan dit belangrijke facet van zijn werk in dergelijke *game*theorieën vaak (al dan niet bewust) voorbij wordt gegaan. Om zijn kritiek te begrijpen, dienen we eerst nog iets verder op *Homo Ludens* in te gaan.

Ten eerste is het belangrijk te begrijpen dat spel niet per definitie begrepen dient te worden als zijnde 'niet-ernst'. Elk spel kan in feite in volledige ernst worden verricht, of het nou kinderspel is, een voetbalwedstrijd of een religieus ritueel. Er zit echter wel een spanning tussen de twee begrippen. Spel kan wel ernstig zijn, maar ergens moet bij de 'spelers' wel het besef blijven bestaan dat het 'slechts' een spel betreft. De latere hoofdstukken van *Homo Ludens* kenmerken zich door een kritische toon. Enerzijds omdat hij waarneemt dat het spelelement meer en meer lijkt te verdwijnen uit onze cultuur. Dit is onwenselijk, daar hij in de voorgaande hoofdstukken heeft betoogd dat spel nodig is om tot culturele ontwikkeling te komen. Dat de cultuur 'meer en meer ernstig [wordt] en […] voor het spel slechts een bijkomstige plaats' (Huizinga, 2008: 103) is ingeruimd, is een ontwikkeling die de cultuur zou kunnen bedreigen. Anderzijds blijkt dat in bepaalde maatschappelijke domeinen (Huizinga wijst onder meer naar de vercommercialisering van de sport) het aanwezige spelelement op zorgwekkende wijze overloopt in ernst. De spanning tussen 'spel' en 'ernst' die in het hele boek al waar te nemen is, lijkt in onze tijd steeds problematischer te worden. Ongezonde competitie, 'aanstellerij' en 'puerillisme' zijn op het eerste gezicht vormen van 'spel' waar bij nader inzien essentiële kenmerken van 'spel' blijken te ontbreken. *Homo Ludens* eindigt met een nogal moralistische noot, waarin Huizinga stelt dat spel altijd een vorm van zelfbeheersing en beperking veronderstelt, een 'zekere vatbaarheid om in je eigen strekkingen niet het uiterste en het hoogste te zien' (Huizinga, 2008: 244). Om 'echte cultuur' te bewerkstelligen dient er bovendien eerlijk te worden gespeeld. Spelbrekers, die in welke vorm dan ook het spel ondermijnen, maken meer kapot dan het spel alleen, zij 'breken' in feite de cultuur.

Op dit punt kunnen we een eerste bruggetje slaan naar de hermeneutiek. Als we ook het culturele verschijnsel hermeneutiek in de geest van Huizinga opvatten als spel, valt op dat ook daarin een notie van 'eerlijk spel' kan worden ontdekt. De kritiek van Gadamer op de reconstructieve hermeneutiek, die onder andere de eenzijdigheid van een dergelijke interpretatie behelst, loopt bij hem uit op een ontwerp van een evenwichtiger en eerlijker 'verstaansspel'. Hiertoe zijn de dingen die Huizinga noemt – zelfbeheersing en zelfbeperking – zeer belangrijk. Immers, om tot een horizonsversmelting te komen zoals Gadamer dit voorstelt is het belangrijk niet enkel de eigen visie uit te dragen maar om ook oog te hebben voor het andere. Het is dus noodzaak om de natuurlijke neiging alles vanuit jezelf te bekijken te doorbreken en ook te proberen buiten de eigen horizon te kijken. Je 'inleven in een andere horizon' heeft iets van een toneel*spel*, waarin de acteurs zich immers ook zo goed mogelijk proberen in te leven in een andere persoon. Op Huizinga's notie van het 'spelbreken' kom ik in het derde deel in het kader van een korte bespreking van Derrida nog terug.

Zoals eerder gesteld is de 'speltheorie' van Huizinga een inspiratiebron en fundament geweest voor latere speldenkers. Voornamelijk bij het ontwikkelen van een 'speltheoretisch kader' voor onder andere de *Game Studies* is Huizinga een veel geciteerde denker. Maar het gaat, zoals ik hierboven al aangaf, in deze gevallen veelal om een al te 'smalle' lezing van *Homo Ludens*, waarbij slechts enkele elementen vrij ruw uit hun oorspronkelijke context zijn getrokken. Vooral Huizinga's poging tot het definiëren van het concept 'spel' in zijn eerste hoofdstuk is veelvuldig gebruikt om een spel te definiëren en het is met name de term '*magic circle*' (waarmee in de Engelse vertaling van *Homo ludens* het Nederlandse 'tooverkring' wordt vertaald) die hierbij vaak wordt aangehaald. Frappant, daar Huizinga dit begrip zelf slechts enkele malen gebruikt en dan specifiek om de heilige plaats van een ritueel aan te wijzen.

De speldefinitie van Huizinga is door veel latere denkers ook verder uitgebouwd en verscherpt. Daarbij verdient met name Roger Caillois (1913-1978) in dit verband te worden genoemd. Zijn boek *Les jeux et les hommes; le masque et le vertige* (1958) is naast Huizinga's *Homo Ludens* een van de belangrijkste inspiratoren van de *game theory* die de afgelopen decennia in de Nieuwe Media wetenschappen tot ontwikkeling is gekomen (Caillois, 2001: 13).

Caillois geeft – terecht – aan dat Huizinga zich in *Homo Ludens* niet waagt aan een eenduidige definitie van 'spel'. Waar Huizinga zich voornamelijk richt op de vraag welke *rol* spel speelt in de cultuur, tracht Caillois een scherpere *definitie* van het begrip 'spel' te ontwikkelen. Hierbij geldt Huizinga's werk wel als een belangrijke inspiratiebron en een rode draad, daar Huizinga in zijn boek talloze spelelementen noemt in culturele uitingen die voorheen nooit zo nadrukkelijk met spel werden verbonden, zoals recht, religie en kunst. Caillois stelt echter dat Huizinga nog niet volledig genoeg is geweest in het omschrijven van de elementen die tot 'spel' kunnen worden gerekend. Om dit alsnog te doen, is hij grondig nagegaan in welke vormen spel zich manifesteert in het dagelijks leven en aan de hand hiervan heeft hij uiteindelijk een schema gemaakt dat '*the entire universe of play*' (Caillois, 2001: 13) in kaart brengt (al geeft hij zelf aan dat ook hier de verdeling nog te kort schiet om het verschijnsel spel geheel recht te doen). Hij verdeelt het begrip 'spel' in enkele kernbegrippen die in een kwadrant te plaatsen zijn. Elke categorie kan op haar beurt weer worden opgedeeld in twee tegengestelde polen.

De categorieën zijn *agon*, *alea*, *mimicry* en *ilinx* en de polen zijn *paidia* en *ludus*. *Paidia*, aan het ene extreme uiterste, kenmerkt zich door verstrooiing, onrust, vrije improvisatie, wispelturigheid en anarchisme (Caillois, 2001: 13). *Ludus*, aan de andere uiterste zijde, tracht deze 'anarchistische onrust' in te binden met eentonige regels, autoritair bevel en kenmerkt zich door inzet, geduld, bekwaamheid en vernuft (Caillois, 2001: 13). Keren we wederom terug naar de hermeneutiek, dan zouden we kunnen stellen dat de reconstructieve en constructieve hermeneutiek zich meer in het *ludus* gedeelte bevinden en dat de deconstructieve hermeneutiek zich als een tegengesteld pool eerder in het *paidia* gedeelte ophoudt. Hoewel op te vatten als twee extremen is het hier echter ook belangrijk op te merken dat beide categorieën zich wel degelijk scharen onder de noemer 'spel'. Derhalve zou ik de stelling willen verdedigen dat in het geval van het spel der hermeneutiek de deconstructieve hermeneutiek zich binnen de hermeneutiek beweegt en niet zozeer een 'anti-hermeneutiek' is.

Binnen de ruimte die deze twee polen beslaan kunnen we volgens Caillois, zoals eerder opgemerkt, vier verschillende categorieën onderscheiden die zich aan beide kanten van de twee polen kunnen bevinden. De eerste categorie is *agon*, het 'concurrerende' spel waarin strijd een vooraanstaande rol inneemt. 'Gelijkheid' is hierbij een belangrijke factor, beide deelnemende partijen dienen bij aanvang op gelijke voet te beginnen, of dit nou betrekking heeft op het aantal deelnemers, de snelheid, de bekwaamheid of iets anders. Een goed voorbeeld hiervan is de schaaksport, waar door middel van klassementen of minder pionnen voor de sterkste speler een evenwichtige speelsituatie wordt gecreëerd (Caillois, 2001: 14). In de praktijk blijkt deze notie van gelijkheid echter vaak moeilijk realiseerbaar. Om bij het voorbeeld van schaken te blijven kan bijvoorbeeld enkel het feit dat een speler als eerste mag beginnen al een onoverkoombaar voordeel zijn. Bij buitensporten kan de zon die in het gezicht van de spelers van een team schijnt een wedstrijd in hun nadeel werken. Zo zijn er diverse voorbeelden te geven. Het doel van *agon* is 'for each player to have his superiority in a given area recognized' (Caillois, 2001: 15). Hiervoor is aandacht, oefening, volhardende toepassing en het verlangen tot winnen

nodig. Wanneer bij kinderen de persoonlijkheid begint te ontwikkelen, treedt reeds het competitieve element in hun spel. 'Kinderlijke' weddenschappen en wedstrijdjes worden vaak gekenmerkt door gewelddadigheid en het opzoeken van pijn. Zo kan er worden gewed wie het langst zijn adem in kan houden, het langst naar de zon kan kijken, het langst niet kan knipperen enzovoorts. De consequenties van verliezen gaan ook vaak gepaard met het krijgen van stompen, klappen of krabben van de tegenstander(s). In dit kinderspel vindt *agon* zijn wortelen en in het 'volwassen' spel vinden we deze vormen van competitie in meer verfijnde gestalte terug (Caillois, 201: 17).

Zouden we de verschillende vormen van hermeneutiek trachten onder te brengen in de indeling van Caillois, dan kunnen we de reconstructieve hermeneutiek van onder andere Schleiermacher en Dilthey onder de noemer *agon* scharen. Het doel van de reconstructieve interpretatie is het 'overwinnen' van de vreemdheid van de te interpreteren uitdrukking (tekst, schilderij, gebouw, handeling etc.). De competitie lijkt echter niet helemaal eerlijk, daar de 'tegenstander', dat wil zeggen het geïnterpreteerde, niet aan zet kan komen. Een 'tekst' heeft immers vaak niet meer de mogelijkheid zich te 'weren' of een discussie aan te gaan. Het gewelddadige element van *agon* zien we dan ook in een dergelijke vorm van interpretatie terug.

De tweede categorie is *alea*, de Latijnse benaming voor het dobbelspel. Deze verzamelterm gebruikt Caillois om alle spelen aan te duiden waar de uitkomst buiten de controle van de spelers ligt. Voorbeelden hiervan zijn bijvoorbeeld dobbelspelen, roulette, kop of munt en de loterij. Waar bij agon wordt geprobeerd toeval zo veel mogelijk te elimineren, zien we dat bij *alea* het juist dit toevallige element is wat deze vorm van spel aantrekkelijk lijkt te maken. De gokspeler moet zich geheel overgeven aan het 'lot'. *Alea* heeft niet als functie om de meest intelligente speler te laten winnen, maar kan juist individuele verschillen geheel doen laten wegvallen waarna spelers als totale gelijken zich over moeten geven aan het vonnis van de kans. *Agon* en *alea* lijken beiden te luisteren naar eenzelfde wet: de 'creation for the players of conditions of pure equality denied them in real life' (Caillois, 2001: 19). Er wordt als het ware een (tijdelijke) nieuwe wereld ontworpen. Alea is samen met *ilinx* te koppelen aan de deconstructieve hermeneutiek van Derrida, hier kom ik zo op terug.

De derde categorie is *mimicry*, het nadoen of zelfs het 'worden' van een fictief karakter. *Mimicry* kan zich op verschillende manieren manifesteren, maar het belangrijkste element is altijd dat de speler doet alsof hij iemand anders is, of zich als een ander voordoet aan anderen. Kinderspel imiteert vaak het volwassen leven, te denken valt aan verkleedpartijtjes, soldaatje spelen met nepgeweren of het spelen van 'vadertje en moedertje'. Onder volwassenen zien we dergelijke verkleedpartijen ook nog wel terug, met als voornaamste voorbeeld het theater. Maar in een meer subtiele vorm zien we *mimicry* ook terug in bijvoorbeeld sportevenementen waar teams dezelfde kleding dragen. Het doel van mimicry is niet om zichzelf en de ander *werkelijk* te doen geloven dat de speler een ander is, maar het draait hier meer om fascinatie en het creëren van een magie (Caillois, 2001: 22). Deze categorie kunnen we verbinden met de constructieve hermeneutiek van onder andere Gadamer. Hoewel we nooit 'werkelijk' in de horizon van een ander kunnen treden, is, zoals ik hierboven reeds opmerkte, bij deze vorm van interpretatie het 'je-kunnen-inleven' in een andere horizon het belangrijkste element. Dit niet met het doel om daadwerkelijk de ander te worden, of een ander hiervan te overtuigen, maar om tot een horizonsversmelting te komen. Dit zou gezien kunnen worden als het creëren van een nieuwe wereld, een magische handeling die iets in het leven roept wat er voor de inleving nog niet was.

Het probleem dat Derrida echter signaleert in de constructieve hermeneutiek tekent zich juist in een vergelijking met het spel goed af. *Mimicry* is immers het *doen alsof*, het blijft 'niet echt'. Ondanks de goede bedoelingen van Gadamer blijft een 'gesprek aangaan' met een te interpreteren 'tekst' door middel van inleving een vorm van spelen. Je kunt nooit echt de ander worden. Zodoende blijft ook een constructieve interpretatie vaak eenzijdig en gaat het net als bij de reconstructieve hermeneutiek nog steeds vaak gepaard met een vorm van geweld. Dat de 'magische bubbel' van het constructieve gesprek gemakkelijk te doorbreken valt, bewees hij wel in zijn 'debat' met Gadamer in 1981 in Parijs, waar hij de dialoog met Gadamer demonstratief uit de weg ging.[1]

Dat de constructieve hermeneutische aanpak in feite een vorm van spel is, wordt door sommige historici ook als zodanig (h)erkend. Zo stelt

mediëvist Laura Kendrick dat (moderne) historici in hun praktische werkwijze in grote lijnen schatplichtig zijn aan de 'speelse' methodologie van Huizinga. Niet alleen de inhoud van bijvoorbeeld *Herfsttij der Middeleeuwen* (1919) en *Homo Ludens* zijn belangrijk voor de wijze waarop veel historici nu te werk gaan, maar ook Huizinga's werkwijze geldt als een voorbeeld. Kendrick:

> '[H]e explained play by himself playing – and in so doing invented the rules of an academic discipline. Huizinga's theorizing about the cultural productivity of play followed upon his study of the European late-fourteenth and fifteenth centuries and his imaginative attempts, in The Waning of the Middle Ages, to understand medieval cultural phenomena, such as chivalry and courtly love' (Kendrick, 2009: 43).

In het licht van de hierboven vastgestelde overeenkomst tussen de reconstructieve hermeneutiek van Gadamer en de *mimicry* van Caillois en de relatie die beider werken hebben met de theorieën van Huizinga, kunnen we enkele interessante dingen vaststellen. Ten eerste merk ik op dat niet alleen Gadamers theoretische opvatting dat het spel de kunst (en in het laatste deel van *Wahrheit und Methode* ook de taal en uiteindelijk de hele historische werkelijkheid) omvat, grotendeels op de gedachtelijnen van Huizinga rust, maar dat hij ook op een meer praktische wijze door hem lijkt te zijn geïnspireerd. We zouden Huizinga's werkwijze met terugwerkende kracht reconstructief hermeneutisch kunnen noemen, die bovendien op grond van mijn van mijn eerdere argumenten als een *vorm van spel*, namelijk *mimicry*, kan worden gezien. Kendrick stelt dat Huizinga middels *imaginative attempts* bepaalde historische elementen tracht te verstaan. Zij spreekt in het kader van latere historici die op eenzelfde 'speelse wijze' naar de geschiedenis kijken ook wel van de verstaanstechniek van de *vision of the smiling eyes*, een manier om 'teksten' te verstaan die reeds sinds lange tijd tot stof zijn vergaan (Kendrick, 2009: 44). Kendrick:

> 'In order to appreciate or understand them, one must imagine oneself into the situation of performance and playing of medieval love poetry or tournaments – an imaginative participation that Huizinga evokes here with his "vision of the smiling eyes." Huizinga's method, like that of certain romantic medievalist before him, was to treat interpretation and understanding of medieval texts or cultural artifacts as a kind of game requiring that modern interpreters accept the absolute alterity of the medieval, setting it within a kind of magic circle, defining it as definitely not our ordinary life' (Kendrick, 2009: 44).

Het behoeft geen uitleg meer dat Gadamer's notie van horizonsversmelting en daarmee zijn constructieve hermeneutiek grote overeenkomsten vertoont met Huizinga's *vision of the smiling eyes*. Het 'anders zijn' van een culturele 'tekst' speelt een grote rol in het werk van veel moderne historici en het 'spel' neemt hiermee een belangrijke positie in binnen de historische hermeneutiek. Veel historici gebruiken de *vision of the smiling eyes* als een middel om de lezer uit te dagen tot de *game of interpretation*, ofwel om hem uit de eigen ervaringshorizon te lokken in de door mij al eerder genoemde nieuwe 'magische ruimte' van de horizonsversmelting. Dit wordt bijvoorbeeld gedaan door historische studies te schrijven in de eerste-persoons-meervoud vorm ('wij'), waarbij de lezer regelmatig wordt aangesproken met 'jij', om hem zo uit te dagen mee te doen met het interpretatiespel dat de historicus speelt (Kendrick, 2009: 47). Dat het hier om *mimicry* blijft gaan en nooit om een werkelijke inleving in bijvoorbeeld het leven van een middeleeuwse man of vrouw, wordt hierbij vaak erkend en benadrukt. Om vrij met Huizinga te spreken: 'de speler is zich er van bewust dat hij speelt'. Dit is echter geen probleem, juist het speelse element zorgt ervoor, zoals Huizinga in *Homo Ludens* betoogt, dat er sprake is van culturele vooruitgang. De horizonsversmelting van Gadamer veronderstelt ook een beweging, waarin wordt geprobeerd de andere horizon in zijn anders-zijn te 'verstaan'. Kendrick: 'having no 'magic cloak', to turn us into medieval interpreters, we need to follow the rules of the game of interpreting like a medieval interpreter, rules that require respect for historical boundaries' (Kendrick, 2009: 48).

We keren terug naar de laatste categorie die Caillois onderscheidt, namelijk *Ilinx*, het soort spelletjes dat draait om duizeligheid en tijdelijke verwarring. Het gaat hier veelal om fysieke activiteiten die de werkelijkheid voor een moment verstoren. Voorbeelden zijn het heel hard om je as draaien, een achtbaan of een hypnose ondergaan. Ook sensaties als hard naar beneden skiën en dansen kunnen worden geschaard onder *ilinx*. Zoals hierboven gesteld, zou de deconstructieve hermeneutiek van Derrida

geplaatst kunnen worden bij *alea* en *ilinx*. *Alea* sluit aan bij het gegeven dat de mens van Derrida bespeeld lijkt te worden, in plaats van dat hij zelf de wereld bespeelt. Bovendien lijkt het (taal)spel dat hij speelt een zelfde soort situatie te creëren als *alea*, namelijk het creëren van een pure gelijkheid tussen de dingen die hen in het echte leven wordt ontzegd. Zoals gesteld gaat zowel constructieve en reconstructieve interpretatie gepaard met 'geweld' doordat er – ondanks goede bedoelingen – vaak op 'oneerlijke voet' wordt gestart. De interpretatie gaat teveel uit van één kant en doet daardoor het geïnterpreteerde vaak weinig recht. De extreme verstrooiing en het 'anarchisme van betekenis' in Derrida's deconstructieve hermeneutiek zorgt er echter voor dat alles op gelijke voet met elkaar 'speelt' (bijvoorbeeld door in *Éperons. De styles de Nietzsche* (1978) de door Nietzsche in de marge van zijn notitieboeken neergeschreven zin 'Ik ben mijn paraplu vergeten' centraal te stellen in zijn interpretatie). Hij zorgt er bovendien voor dat, net zoals bij *ilinx*, de interpreet voor een moment uit zijn *comfort zone* – zijn ervaringshorizon – wordt geslingerd om verdwaasd in de chaos van betekenis die Derrida achterlaat rond te dolen. In deel drie van deze paper kom ik terug op de rol van Derrida als 'spelbreker'. Voordat ik mij hiertoe begeef, ga ik eerst nog verder in op de rol die Gadamer het spel toekent in de hermeneutiek.

Zoals eerder al gesteld, leunt Gadamer in zijn opvatting van het spel als een 'alomvattend gegeven' in grote mate op de speltheorie van Huizinga. Dit komt vooral tot uiting in Gadamer's opvatting van spel als '*the essence of things*' (Vikhagen, 2009: 2). In navolging van de kritiek van Heidegger op het subjectdenken (in de cartesiaans-kantiaanse traditie) bekritiseert Gadamer in zijn boek *Wahrheit und Methode* de 'subjectivering' van het 'spel', dat wil zeggen de opvatting dat de menselijke speler als subject het spel bepaalt. Bij Gadamer lijkt echter het spel *zelf* tot 'subject' van het spel te worden. Hij stelt de 'geestesstaat' en houding van degene die speelt ondergeschikt aan het spel, dat wil zeggen aan de 'speelse houding' van een kunstwerk (in latere hoofdstukken van *Wahrheit und Methode* en latere werken lijkt het begrip spel zich verder uit te breiden als alomvattend fenomeen, niet enkel meer binnen de kunsten). De rol van de speler hierin, is niet die van een vormend en bepalend subject, maar veeleer die van een katalysator van het spel. Gadamer maakt dus een onderscheid tussen het spel zelf en het 'spelgedrag' van de speler, dat hij schaart onder andere subjectieve vormen van gedrag (Gadamer, 1986: 107).

De spanning tussen 'spel' en 'ernst' zoals waargenomen in het werk van Huizinga is ook aanwezig in Gadamers speltheorie. Hij stelt dat voor de speler het spel niet 'serieus' is, omdat hij 'speelt'. Het spel zelf wordt echter gekenmerkt door een '*heilige Ernst*' (Gadamer, 1986: 107). De notie van 'heilige ernst' komt ook regelmatig terug bij Huizinga. Spel (en daarmee ook de dieperliggende 'ernst' ervan) kan enkel worden bereikt wanneer de speler zichzelf verliest in het spel, hoewel hij zich er wel van bewust blijft dat het 'maar' een spel is. Er lijken zich in de speler dus gelijktijdig twee tegengestelde houdingen af te spelen. Het 'weten dat men speelt' is daarmee een enigszins paradoxaal gegeven en in het weten van de speler dat hij speelt 'weet hij niet precies wat hij "weet" door dit weten (Gadamer, 1986: 108)'. Precies daarom kan het antwoord op de vraag wat spel precies is niet in het 'onwetend-wetende subject' worden gelegd. Om deze reden richt Gadamer de focus op het 'speels zijn' van de dingen, meer in het bijzonder op de speelsheid van de kunst.

Het kunstobject is 'kein Gegenstand […], der dem für sich seienden Subjekt gegenüber steht. Das Kunstwerk hat viel mehr sein eigentliches Sein darin, daß es zur Erfahrung wird, die den Erfahrenden verwandelt' (Gadamer, 1986: 108). Het 'subject' van de kunstervaring ligt hierdoor niet bij de subjectiviteit van de ervarende persoon of bij het kunstobject zelf, maar in de omvattende ervaring waarbinnen het kunstwerk en de aanschouwer zich beiden bevinden. De '*Seinsweise des Spieles*' strekt zich daarmee verder uit dan enkel de horizon van de mens, maar bestaat ook wanneer er geen 'speelse subjecten' aanwezig zijn. Deze 'speelse subjecten' zijn eerder één van mogelijke wijzen waarop spel tot uitdrukking ('*Darstellung*') kan komen. Het Duitse '*Spiel*' betekent van origine 'dans'. Dit element van beweging speelt een belangrijke rol in Gadamers opvatting van spel. De beweging van spel heeft geen doel buiten zichzelf, het vernieuwt zichzelf continue. Wie of wat er speelt is hierbij niet van belang, evenmin of er een menselijk subject is dat het speelt of niet. Het spel manifesteert zich als het ware in de beweging als zodanig (Gadamer, 1986: 108). 'Damit Spiel sei, muß zwar nicht ein anderer wirklich mitspielen, aber es muß immer ein anderes da sein, mit dem der Spielende spielt und das dem Zug des Spielers von sich aus mit einem Gegenzug antwortet'

(Gadamer, 1986: 111).

Een ander belangrijk kenmerk van spel, dat zich ook min of meer expliciet manifesteert in *Homo ludens*, is het gegeven dat spelen ook altijd een 'gespeeld worden' impliceert. Spel staat nooit geheel onder controle van de speler en het is groter dan de speelse handeling alleen. Degene die speelt is belangrijk voor de manifestatie van spel, maar wordt na aanvang van het spel ook vooral zelf gespeeld. Hieruit blijkt wederom dat het werkelijke subject van het 'spel' niet de speler is, maar in feite het spel zelf. Deze gedachte breekt met het Cartesiaanse subject, daar dit betekent dat het menselijk subject geen volstrekte autonomie heeft. Hoewel de mens in vele vormen van spel wel een belangrijke 'katalysator' is en 'speelse' dingen als kunst, theater en taal heeft geschapen, is hij niet heer en meester over deze vormen van spel en wordt zelf opgenomen in het geheel. Deze gedachte, zo zullen we in het volgende deel zien, speelt ook in het werk van Derrida een grote rol. Dat Gadamer echter, enigszins misleidend, alsnog blijft spreken van het spel als zijnde het 'subject', wordt hem door Derrida niet in dank afgenomen. In zekere zin is het nodig dat Gadamer blijft spreken in voor ons bekende termen, anders is er letterlijk geen beginnen aan het bedenken van een nieuw verstaan. Derrida's deconstructie brengt echter de problemen van dergelijke 'taaltradities' aan het licht en breekt met conventies. Of speelt ook hij (het spel mee)…?

Spelbreker

Hoewel Gadamer al een moedige poging doet te ontsnappen aan het 'subjectdenken', lijkt hij er toch niet onder uit te komen van een níeuw subject te spreken. Dat is niet vreemd, daar volgens de hermeneuticus Samuel IJseling gesteld kan worden dat:

> '[e]lke taaldaad, van welke aard die ook is, een bepaald kader [veronderstelt], een context die zowel talig als reëel is. […] Zoals men zich geen enkel voorwerp en geen enkele gebeurtenis kan voorstellen of denken zonder horizon, zonder een tijdruimtelijke wereld en zonder andere voorwerpen of gebeurtenissen waarnaar verwezen wordt en waarop een voorwerp of gebeurtenis aangewezen is om er te zijn of te gebeuren, zo kan men zich evenmin een woord, zinsnede of tekst voorstellen of denken zonder dat deze op één of andere manier verwijst naar een contekst en daarop aangewezen is om te kunnen worden gesproken of geschreven, gehoord of gelezen' (IJseling, 1990: 9).

Met andere woorden, de speltheorie van Gadamer heeft enkel een bepaalde betekenis wanneer hij in een bepaalde context wordt geplaatst. Bij Gadamer is dat de 'subjecttraditie', waar hij zich tegen afzet, maar die juist daardoor ook de context vormt die het mogelijk maakt Gadamers inzet te begrijpen. Derrida trekt echter de gehele notie van 'context' in twijfel. Hij stelt dat een context voortdurend en tot in het oneindige naar alle kanten kan worden uitgebreid en dat daarmee het verschil tussen tekst en context onbeslisbaar is. De 'grenzen' die wij aan een context leggen zijn daarmee altijd willekeurig en kunnen bovendien worden verlegd. Hier komt wederom de notie naar voren die ik hierboven aanhaalde: er wordt iets beslist dat in feite 'onbeslisbaar' is. Wanneer een woord uit een context wordt gehaald kan het bovendien weer op een andere manier gaan functioneren en nieuwe betekenissen creëren. Deze 'herhaling' vindt niet alleen plaats wanneer ik bijvoorbeeld een citaat van IJseling opneem in mijn betoog en daarmee zijn woorden uit de oorspronkelijke context haal, maar ook al wanneer ik zijn woorden lees. Ik herhaal dan als het ware iets wat al is geschreven maar geef er nieuwe betekenis aan doordat ik er vanuit een bepaalde context (in dit geval mijn analyse van het 'spel') naar kijk. Derrida spreekt in dit verband ook wel van enten (IJseling, 1990: 10), een verwijzing naar het planten van stekjes van een boom of plant op een nieuwe plaats. Zoals dit kan worden gedaan met stekjes, kan ook een zinsnede of woord op een andere tekst of tekstueel veld worden geënt. Een zeer belangrijke opmerking voor de verloop van dit betoog is dat het volgens Derrida wel zo is dat elk 'woordstekje' op de een of andere manier op andere teksten geënt is, een tekst helemaal op zichzelf, zonder enige context, dat bestaat gewoonweg niet. Ook hierin is weer een parallel met het spel te vinden. Gadamer stelt:

> 'Das menschliche Spiel verlangt seinen Spielplatz. Die Abgrenzung des Spielfeldes – ganz wie die des heiligen Bezirks, wie Huizinga mit recht betont – setzt die Spielwelt als eine geschlossene Welt der Welt der Zwecke ohne Übergang und Vermittlungen entgegen' (Gadamer, 1986, 113).

Dit citaat is mooi om twee redenen. Enerzijds zou je kunnen stellen dat het in lijn ligt met Derrida's opvatting van woorden en zinsneden die als het ware spelen in een tekstueel veld. Anderzijds laat het citaat van Gadamer precies het problematische van het spel van interpretatie zien dat Derrida voor ogen heeft. Interpreteren doen we inderdaad altijd in een context, een speelveld, een 'tekstuele akker', of hoe je het ook wilt noemen. Dat is nodig omdat, zoals ook Derrida stelt, er altijd een context nodig is om betekenis te bewerkstelligen. Er zijn bepaalde regels nodig en natuurlijk de verschillende spelers. Echter, en daar ligt het problematische in het citaat, zoals spel nooit helemaal los staat van de wereld, zo staat een interpretatie ook nooit alleen. Het tekstuele veld verandert, er komen nieuwe 'entjes' in, sommige interpreten zien entjes groeien die een ander eerder niet zag. Betekenisproductie en ook het verstaan laten zich daarom nooit volledig controleren en beheersen. En het lijkt niet toevallig dat ook Derrida in dat verband het woord 'spel' gebruikt om die ongrijpbare dimensie van het verstaan te benoemen. In *L'écriture et la difference* – in het hoofdstuk waarin het woord 'spel' reeds in de titel opduikt: 'La structure, le signe et le jeu dans le discours des sciences humaines' – stelt hij, sprekend over de onmogelijkheid om de totaliteit van een bepaalde zaak uit te drukken, dat dit niet zozeer komt – zoals de klassieke hermeneutiek geneigd is te denken – doordat het onderwerp meer bevat dan de eindige mens tot uitdrukking kan brengen ('Il y a trop et plus qu'on ne peut dire'), maar omdat het verstaan wordt meegesleept in een onbeheersbaar spel:

> 'Mais on peut determiner autrement la non-totalisation: non plus sous le concept de finitude comme assignation à l'empiricité mais sous le concept de jeu. Si la totalisation alors n'a plus de sens, ce n'est pas parce que l'infinité d'un champ ne peut être couverte par un regard ou un discours finis, mais parce que la nature du champ — à savoir le langage et un langage fini — exclut la totalisation: ce champ est en effet celui d'un jeu, c'est-à-dire de substitutions infinies dans la clôture d'un ensemble fini. Ce champ ne permet ces substitutions infinies que parce qu'il est fini, c'est-à-dire parce qu'au lieu d'être un champ inépuisable, comme dans l'hypothèse classique, au lieu d'être trop grand, il lui manque quelque chose, à savoir un centre qui arrête et fonde le jeu des substitutions' (Derrida, 1967, 423).

Een tekst – of zoals Derrida het noemt een 'akker' – is dus wel eindig, maar het 'spel' regeert hier (een onbegrensd interpreteren), omdat de 'tekstuele akker' een 'tekstueel anker' mist en daardoor nooit (definitief) vastgelegd kan worden. Dat de teksten van Derrida's eigen hand zeer weloverwogen zijn geschreven en zijn gebruikte termen vaak welhaast 'kapot geanalyseerd' worden is dan ook geen toevalligheid. 'Levinas heeft ooit van Derrida gezegd dat hij onder (bijna) elk woord een bom, een mijn geplaatst heeft waardoor het levensgevaarlijk wordt deze woorden nog te gebruiken zonder uiterste voorzichtigheid en behoedzaamheid' (IJseling 1990: 11). IJseling noemt Derrida's werkwijze dan ook wel de 'techniek van de vertraging', een wijze waarmee Derrida zijn tekstuele veld als het ware aftast alsof het geheel in duister is gehuld (IJseling, 1990: 11). Deze 'onthaasting' staat in scherp contrast met de gehaastheid die het alledaagse leven – en zeker het alledaagse leven in onze moderne, snelle leefwereld met zijn overdaad aan prikkels – kenmerkt. Als eindige wezens kunnen we immers niet oneindig lang blijven stilstaan in het alledaagse verstaan. Het 'niet-denken' en 'niet-zeggen' is daarmee een onderdeel van het mens-zijn (IJseling, 1990: 13). Al het interpreteren is volgens Derrida een 'is-denken', ofwel het vellen van een oordeel. Maar zo is het leven nu eenmaal, de mens staat als het ware onder 'een dubbele wet' (IJseling, 1990: 14). Enerzijds kunnen en mogen we eigenlijk niet interpreteren omdat we dan altijd een oordeel vellen over iets 'onbeslisbaars'. Anderzijds moeten we wel! Ook bij klassieke hermeneutici als Dilthey treffen we dat inzicht reeds aan. De Mul:

> 'Dilthey merkt […] herhaaldelijk op dat het de mens onmogelijk is om voortdurend te leven in het bewustzijn van de relativiteit van onze meest fundamentele uitgangspunten. Wanneer we dit besef voortdurend tot ons zouden toelaten zou ieder handelen en ieder denken worden verlamd. […] De mens [moet] om te kunnen leven met zijn fundamentele eindigheid wel denken en handelen alsof hij eeuwige waarheden en waarden schept' (De Mul, 1993: 384).

Dit citaat verwijst wederom naar het mimetische karakter van de hermeneutiek. Tevens laat het de spanning zien die ook Derrida veronderstelt; we moeten en mogen eigenlijk niet interpreteren, maar anderzijds kunnen

we niet anders. Het is dan ook niet verwonderlijk dat het boek waarin Derrida een poging doet tot een radicaal nieuw filosofisch denken omtrent interpreteren te komen *Marges de la philosophie* (1972) heet. Net zoals hij Gadamer verwijt in de valkuilen van de filosofische tradities te vallen, lijkt hij zich niet over de 'rand' van de filosofie te kunnen begeven, maar balanceert hij er vervaarlijk op. Met zijn techniek van traagheid en zijn destructieve tekstanalyse legt hij regels bloot waarvan zijn filosofische 'medespelers' (– het is wellicht geen toeval dat de heren van *Monty Python* in een van hun sketches de grootste denkers van de wereld in toga en al op een voetbalveld plaatsten –) zich wellicht niet meer bewust zijn. Hij creëert letterlijk en figuurlijk een afstand van het speelveld, en kan daarom worden gezien als een – ietwat opstandige – commentator die het gebeuren van bovenaf aanschouwt en ontrafelt totdat er van het spel ogenschijnlijk weinig meer overblijft. Weer terugkoppelend naar de begrippen *ludus* en *paida* van Caillois kan mijns inziens worden gesteld dat dit niet betekent dat Derrida zich buiten het speelveld van de hermeneutiek zou bevinden. De deconstructieve hermeneutiek zou zoals eerder gesteld onder *paidia* kunnen worden geschaard en *ludus* en *paida*, zo stelt Vikhagen, vormen samen 'a unity of order and disorder' (Vikhagen, 2009: 2). Zoals verstaan en niet-verstaan onlosmakelijk en paradoxaal bij elkaar horen, zo verhoudt de destructieve, anarchistische, wanorde creërende hermeneutiek van Derrida (*paidia*) zich tot de andere, regelgebonden en overzichtelijkere vormen van hermeneutiek (*ludus*).

Huizinga stelt dat de spelbreker, degene die de regels ondermijnt, die het spel blootlegt en tegelijkertijd ontbindt, ook de cultuur ondermijnt. In zekere zin is Derrida te zien als een spelbreker. Tijdens het 'debat' in 1981 in Parijs, waar hij weigerde Gadamers aanbod om in een constructieve dialoog te treden aan te nemen, is hij letterlijk een spelbreker; hij speelt het 'gesprekspel' van Gadamer niet mee.[2] Echter, uit zijn eigen uitspraak dat een woord of zinsdeel nooit alleen kan staan, maar altijd een context nodig heeft, blijkt al dat ook zijn radicale, destructieve hermeneutische taalspel niet los kan staan van de rest. Sterker nog, zonder geschreven woorden en het 'spel' dat de andere hermeneutici zo braaf volgens de regels volgen, kan zijn destructieve hermeneutiek niet eens bestaan.[3]

In computer*games* zitten regelmatig *bugs* die maken dat de speler van het spel kan breken met de regels van het spel zonder dat dit de intentie van de makers is geweest. Sommige spelers maken er zelf een spelletje van om zo veel mogelijk fouten in een spel te ontdekken en de game volgens nieuwe regels uit te spelen. Regelmatig leidt het creatieve gebruik van de bugs tot aanpassingen in de volgende game in de reeks en worden de 'fout' en het niet volgens de regels spelen onderdeel van het spel. Een voorbeeld hiervan is bijvoorbeeld de *hammer jump* in *shooter games*. Vikhagen:

> 'Usually, and because of the impossible task of removing all programming bugs in a computer game, the ability to break rules is not intentional from the game developers. Such is the case for instance with the "hammer jump". When players in first person shooter games discovered they could shoot into the ground while jumping at the same time, they often used this initially unintended feature to reach places where they could be safer or in other ways gain advantages, even though they would take damage while jumping. This feature turned out to be popular, and was included in later games, this time as an intended feature' (Vikhagen, 2009: 4).

Het is hier, in de grijze zone tussen bug en *intended feature*, tussen *ludus* en *paida*, tussen orde en chaos, waar we Derrida kunnen plaatsen, in de marges van het verstaansspel. Zoals ik met Vikhagen laat zien, betekent dit niet zozeer dat Derrida een spelbreker is, maar dat hij zich op de rand begeeft, een moment balanceert en uiteindelijk toch weer deel wordt van het spel. Spel, zo stelt Huizinga is culturele vooruitgang. In feite is Derrida geen 'spelbreker' in de negatieve zin die Huizinga voor ogen heeft, maar zorgt hij met zijn destructieve aanpak voor meer – letterlijke en figuurlijke – speelruimte. Ook hij, in zijn rol van 'spelbreker', wordt uiteindelijk weer deel van het spel, al is het maar omdat ik zijn woorden nu herlees en hergebruik om tot dit betoog te komen.

Conclusie

In het voorafgaande heb ik getracht te illustreren dat het concept 'spel' in meerdere opzichten met de hermeneutiek verbonden is en dat het zien van de hermeneutiek als een spel verhelderend kan werken. Ik heb laten zien dat Derrida's deconstructieve hermeneutiek niet als een 'anti-herme-

neutiek' moet worden gezien, maar dat ook hij onderdeel uitmaakt van het 'verstaansspel'. Om dit te verhelderen heb ik laten zien hoe de deconstructieve hermeneutiek in het licht van Caillois' speltheorie op te vatten is als *paidia*, waar de constructieve- en reconstructieve hermeneutiek eerder als *ludus* te zien zijn; beiden vallen onder het 'spel'. De vergelijking tussen Caillois' vier vormen van spel – *agon*, *alea*, *ilinx* en *mimicry* – laat bovendien zien dat de verschillende behandelden hermeneutici ieder een eigen vorm van spel lijken te spelen, met eigen kenmerken en regels. Derrida begeeft zich daarbij regelmatig op de rand, maar kan mijns inziens niet worden gezien als een spelbreker in de negatieve zin zoals we die bij Huizinga vinden. Eerder creëert hij meer speelruimte voor het aanpassen en misschien zelfs wel veranderen van bestaande 'regels'. Tevens wil ik toevoegen dat het geenszins zo is dat een vergelijking van de hermeneutiek als spel zou impliceren dat de hermeneutiek een 'onserieuze' bezigheid zou zijn. In lijn van Huizinga kan worden gesteld dat juist spel een cultuurscheppende aangelegenheid is. Dat elke interpretatie onvermijdelijk een oordeel is en bovendien geen eeuwige waarheid, betekent niet dat we het dan maar zouden moeten laten. Dat kunnen we niet, zo hebben we al gezien, maar dat moeten we ook niet willen. Een totaal gedeconstrueerde werkelijkheid zou er een van extreme fragmentatie zijn waarin tussen de rondvliegende kluitjes tekstueel veld enkel nog de felrood knipperende woorden '*game over*' vallen te ontwaren.

Elize de Mul (1987) heeft een bachelorgraad in de Theater- Film- Televisie en Nieuwe Mediawetenschappen (2010) en volgt momenteel de master Nieuwe Media en Digitale Cultuur (Universiteit Utrecht) en de master Filosofie van de Geesteswetenschappen (Erasmus Universiteit Rotterdam).

'Het spel der betekenissen' is ter afronding van het mastervak 'Filosofie van de geesteswetenschappen: Actuele thema's in de hermeneutiek' van prof. dr. Jos de Mul geschreven en nagekeken en voor publicatie in het ESJP genomineerd door dr. Awee Prins.

Literatuur

Caillois, R. (2001) *Man, Play and Games*. Urbana and Chicago: University of Illinois Press.

De Mul, J. (1993) *De Tragedie van de Eindigheid*. Kampen: Kok Agora.

De Mul, J. (in druk) 'Horizons of hermeneutics: Intercultural hermeneutics in a globalizing world'. In: *Frontiers of Philosophy in China*, Vol.2, 2011.

Derrida, J. (1967) *L'écriture et la difference*. Paris: Éditions du Seuil.

Derrida, J. (1978) *Éperons: Les styles de Nietzsche*. Paris: Flammarion.

Derrida, J. (1989) *Marges van de filosofie*. Hilversum: Gooi&Sticht.

Gadamer, H-G. (1986) *Wahrheit und Methode: Grundzüge einer philosophische Hermeneutik*. Gesammelte Werke. Band 1. Tübingen.

Huizinga, J. (2008) *Homo Ludens: Proeve eener bepaling van het spel-element der cultuur*. Amsterdam: Amsterdam University Press.

IJsseling, S. (1999) 'Jacques Derrida: een strategie van de vertraging'. In: Widdershoven, G. A. M., De Boer. T. (red.) *Hermeneutiek in Discussie*. Delft: Uberon: 16-22.

Kendrick, L. (2009) 'Games Medievalists Play: How to Make Earnest of Game and Still Enjoy It'. In: *New Literary History*, Vol. 40: 43-61.

Michelfelder, D.P. en Richard E. Palmer (1989) *Dialogue and deconstruction: the Gadamer-Derrida encounter*. Albany: State University of New York Press.

Vikhagen, A. K. (2009) 'Gadamer's concept of play'. URL: http://sensuousknowledge.org/wp-content/uploads/2009/04/sk1_ak_vikhagen_spiel.pdf.

Notes

1 Zie Diane P. Michelfelder en Richard E. Palmer. (1989). *Dialogue and deconstruction*: *The Gadamer-Derrida encounter*, *SUNY series in contemporary continental philosophy*. Albany: State University of New York Press.

2 Zie Diane P. Michelfelder en Richard E. Palmer. (1989). *Dialogue and deconstruction : the Gadamer-Derrida encounter*, SUNY series in contemporary continental philosophy.

3 In het eerder genoemde 'La structure, le signe et le jeu dans le discours des sciences humaines' merkt Derrida met betrekking tot (het spel van) de metafysica op: "[…] *nous ne pouvons énoncer aucune proposition destructrice qui n'ait déjà dû se glisser dans la forme, dans la logique et les postulations implicites de cela même qu'elle voudrait contester*" (Derrida, 1967, 412).

ESJP
#1 | 2011

The Argument for Anomalous Monism, Again

Deren Olgun

1. Introduction

The main focus of the contemporary debate on mental causation has centred on whether mental events can cause other events in virtue of their mental properties, or only in virtue of their physical ones. Whilst reductive physicalists maintain that the only properties that exist are physical properties, and that any mental or other "higher-level" properties are only properties in virtue of their being identical with some physical property, non-reductive physicalists maintain that mental or other "higher-level" properties are irreducible to physical ones. A common charge that has been levelled against non-reductive physicalists is that if mental properties are irreducible then they must be causally inert or "epiphenomenal" since it cannot be the case that both mental and physical properties are simultaneously causal, a position which the literature has come to call *property epiphenomenalism*. This paper argues that the charge of property epiphenomenalism is misplaced when it is applied to Donald Davidson's anomalous monism. Davidson cannot be accused of property epiphenomenalism because properties do not feature in his ontology and, therefore, play no role in his account of causal relations between events. Davidson's work and name have become embroiled in the debate about property epiphenomenalism because he is mistakenly thought to be working in an old tradition that accepts properties as an ontological category and that maintains that it is only in virtue of their properties that events have the effects that they do. Whilst the conceptual irreducibility of mental types to physical types (predicate dualism), which is the hallmark of non-reductive physicalism, is, for non-reductive physicalists, a consequence of an ontological non-reductivism (property dualism), for Davidson the start and end point is the denial of conceptual reduction (i.e. the acceptance of predicate dualism only). Thus Davidson is neither a non-reductive physicalist, nor can he be accused of property epiphenomenalism – properties simply do not enter into his philosophical scheme.

This paper is divided into eight sections. The first section introduces the mental causation debate by giving a brief outline of its history up to the contemporary debate about property epiphenomenalism. The second section outlines the argument for anomalous monism. The third section outlines one of the earliest arguments that asserted that anomalous monism implies property epiphenomenalism. The fourth section outlines various non-reductive physicalist responses that attempt to defend non-reductivism. The fifth section outlines Kim's influential "overdetermination" argument – the claim that non-reductive physicalists must accept either property epiphenomenalism or overdetermination. The sixth section makes the case that anomalous monism cannot imply property epiphenomenalism. The seventh section deals with Sophie Gibb's (2006) criticisms of Davidson's underlying approach to ontology, as instances of more appropriate criticisms. The final section concludes.

2. The Mental Causation Debate

Cartesian dualism is the classic form of *substance dualism*. Descartes maintained that there are two kinds of substance, material and mental, and that man is a union of a spatially extended, material substance (the body), which is incapable of thought or feeling, and a spatially un-extended, mental substance (the mind or *soul*), which thinks and feels.

Now, pre-theoretic conceptions of agency hold that the mind and the

body causally interact. When, for example, Alfred summons the waiter (by raising his hand, say), we say that his wanting to order caused him to do so, just as we say that my intention to read Emile Zola's *Germinal* causes my buying of the text. And, in agreement with these common sense intuitions, Descartes also maintained that the mind and the body causally interact. However, one of the main problems for Cartesian dualism is how to marry the possibility of mental-physical causal interaction with the total independence of mental and material substances – it is very difficult to see how an un-extended, immaterial substance with no presence in physical space could causally influence material bodies that are subject to the laws of physics. And indeed, as Kim notes, the 'inability to explain the possibility of "mental causation", how mentality can make a causal difference to the world, doomed Cartesian dualism' (Kim, 1996: 4).

The majority position (see Kim, 1996; Crane 2003) in contemporary philosophy of mind rejects Cartesian dualism in favour of a kind of *monism*, which argues that there is only one kind of substance. More specifically, it is a physicalist or "materialist" monism that has come to dominate the debate, a position that is generally known as *physicalism*. Kim defines *ontological physicalism* as the position that 'there are no concrete existents, or substances, in the spacetime world other than material particles and their aggregates' (Kim, 1996: 211) and observes that 'in most contemporary debates, ontological physicalism forms the starting point of discussion rather than a conclusion that needs to be established.' (*Ibid.*: 211)

However, embracing physicalism does not resolve the problem of mental causation that dogged substance dualism: it still remains unclear as to how mental events or objects (whatever their apparent relation to the physical substance that constitutes them) can causally interact with physical events or objects. The debate has simply shifted its focus from substances to properties, as Kim notes 'the most intensely debated issue – in fact, the only substantive remaining issue – concerning the mind-body relation has centred on *properties* – that is, the question *how mental and physical properties are related to each other*' (*Ibid.*: 211-212 – emphasis in original). To make sense of the debate to which Kim refers it will be helpful to draw on the distinction between *token identity* and *type identity.* Token identity states that any mental event or object is identical with some physical event or object – accepting token identity implies a rejection of substance dualism. Type identity, in contrast, holds that mental event types are identical with/reducible to physical event types – "types" here are generally taken to mean properties, although, as we will later observe, it can also mean predicates.

Accepting both token identity and type identity is a *reductive physicalist* position. Reductive physicalists deny mental causation because, as they maintain, a mental event only has the causal power it does in virtue of its being identical with some physical event; that it is only because the event is of a certain physical type that it is the cause it is, and its being of a certain mental type is just a consequence of its being that physical type.

In contrast to the reductive physicalist position, Donald Davidson's *anomalous monism* (1970, 1993) seeks to defend the possibility of mental causation. The idea that reasons are causes of actions is a central part of Davidson's philosophy (see Davidson, 1963) and anomalous monism represents a stance in the philosophy of mind that combines his position on mental causation with other elements of his philosophy, including his views on events (see Davidson, 1969), causation (see Davidson, 1967b) and semantics (see Davidson, 1967a, 1974a, 1977). Whilst anomalous monism accepts token identity, it denies type identity (where by "types" Davidson would have in mind predicates rather than properties), holding that mental types are irreducible to physical ones.

In response, many critics (e.g., Honderich, 1982; McLaughlin, 1993; Kim 1993a) have argued that anomalous monism is inherently contradictory and that the only reasonable way to resolve the contradiction is to accept that mental events are epiphenomenal, or rather that *mental properties* are epiphenomenal – the general argument taking the form that it is only in virtue of an event's being a certain physical type that it has the effect that it does, and not in virtue of its being a certain mental type, that mental events do not cause anything *qua* mental; a position which we shall call *property epiphenomenalism*. In response to these criticisms, arguments in defence of anomalous monism (e.g., LePore & Loewer (1987), Macdonald & Macdonald (1991), Macdonald (2007)), which are usually dubbed *non-reductive physicalism*, have been proposed. These arguments,

in general, contort the positions developed by Davidson's critics in such a way as to (attempt to) restore causal efficacy and relevance to the mental properties of events.

In so far as the debate is one between reductive and non-reductive physicalists, or one between non-reductive physicalists of different stripes, asking whether or not their positions imply property epiphenomenalism is a reasonable line of inquiry. Where it ceases to become reasonable is when Davidson's anomalous monism is made the object of this line of attack either explicitly or because it has been classed, as it frequently is (e.g. Jacob, 2002), as a form of non-reductive physicalism. As Davidson (1967b, 1980, 1993) maintains, and as is reiterated by Tim Crane (1995) and Sophie Gibb (2006), causal relations are extensional relations between events; they are independent of the manner of their description. To ask, within the framework of anomalous monism, whether a mental event causes a physical event *in virtue* of its mental properties or only *in virtue* of its physical properties is to misinterpret a fundamental tenet of the theory of anomalous monism. Properties do not feature in the ontological system on which anomalous monism is based, owing to Davidson's holistic, truth-conditional approach to semantics and metaphysics, and, as such, it simply cannot be criticised for rendering properties epiphenomenal. Properties are simply irrelevant to anomalous monism. The following sections make that case.

3. Anomalous Monism

Davidson (1970) originally presented anomalous monism as a solution to an apparent paradox between three principles which he was inclined to accept:

1. The Principle of Causal Interaction (CI): 'That at least some mental events interact causally with physical events' (Davidson, 1980: 208)

2. The Nomological Character of Causality (NCC): All causal relations instantiate a strict law

3. The Anomalism of the Mental (AOM): There are no strict psychophysical laws

In terms of the discussion of previous section, CI is just the acceptance of the possibility of what we have been calling mental causation. NCC is, by Davidson's own acknowledgment (Davidson, 1970: 209), unsupported. AOM is defended by appealing to Quine's argument of the indeterminacy of translation: it is not possible to formulate strict laws that relate the two, distinct, conceptual realms, because mental concepts are not explicable in physical vocabulary, nor physical concepts in mental vocabulary. As Davidson notes:

> 'There are no strict psychophysical laws because of the disparate commitments of the mental and physical schemes' (Davidson, 1970: 222).

It is from this principle that Davidson's denial of *type identity* arises; because there is no systematic correlation between mental and physical types there is no basis for reduction from the mental to the physical.

So, in anomalous monism (1) the mental interacts causally with the physical, (2) any given causal interaction is describable by a strict law, and (3) there are no psychophysical strict laws. Since Davidson holds that the physical is *causally closed*[1], i.e., that "any physical effect must have a sufficient physical cause" (Crane, 1995: 7), he maintains that the strict law instantiated by any causal relation is always a strict physical law. Therefore, any causal relation between two events is describable by a strict physical law, including the interaction between a mental event and a physical event. Davidson solves the apparent paradox by concluding that mental events *just are* physical events; he endorses *token identity*. However, an event is mental only insofar as it is given a mental description, and is physical only insofar as it is given a physical description – the two are both descriptions of the same event in different vocabularies. Hence, unlike physicalists, Davidson gives no ontological primacy to physical descriptions but, to put it loosely, regards them only as a kind of description that – in virtue of the formation of our physical concepts – *happens* to allow for the statement of strict laws. As a result, Davidson's position constitutes a kind of neutral monism, not, strictly speaking, a kind of physicalism.

4. Causation, *In Virtue Of*

Ted Honderich (1982) was one of the first to critique anomalous monism for rendering mental properties epiphenomenal. Honderich argues that only those properties of an event that can enter into lawlike connections could be causally relevant and since, in anomalous monism, the mental properties of an event cannot enter into lawlike connections, he argues that it renders mental properties epiphenomenal; that anomalous monism entails *property epiphenomenalism*, or, as it is otherwise called (e.g. McLaughlin, 1993), *type epiphenomenalism*.

Honderich's argument begins by noting Davidson's ontological conception of an event as an "irreducible entity" and adds to this position the claim that 'an event has an indefinite number of properties, features or aspects' (Honderich, 1982: 60). To illustrate this point, consider a brick: it could be said to have, amongst others, the properties of "hardness", "redness", "coarseness" and "being cuboid". Given that events have properties in this same way, Honderich argues that 'it is in virtue of certain of its properties rather than others that an event is the cause it is' (*Ibid.*: 61). So, in the event of the brick breaking a window, it is the brick's property of hardness coupled with the window's property of fragility and a small set of other properties (e.g. the brick's velocity) that break the window. The two properties are thus relevant to the cause and effect relation, whilst properties such as the "redness" or "coarseness" of the brick are irrelevant to it. So, Honderich concludes:

> 'If the ground for saying that two events are in lawlike connection is that they are cause and effect,[2] and it is the case that all of their properties save some residue are irrelevant to their being cause and effect, then they are in the given lawlike connection solely in virtue of that residue of properties' (*Ibid.*: 62).

Honderich thus claims that causal relations exist only between certain properties of events, and that a lawlike connection exists in virtue of these properties. He calls this claim the "Principle of the Nomological Character of Causally Relevant Properties".

Now, since AOM implies that mental properties cannot enter into strict lawlike connections with physical properties, Honderich contends that they are not captured by this Principle, i.e., mental properties are not causally relevant. Therefore, either we must reject AOM or reject the claim that mental events interact with physical invents (CI). Honderich suggests that we should reject the strong form of CI, and argues, instead, that it is mental events *as* physical events that cause physical events. So, whilst mental events do interact with physical events, it is only in virtue of their being identical with physical events. Therefore, since the mental properties of mental events are epiphenomenal in the causal relation, Davidson must be a property epiphenomenalist.

This accusation of property epiphenomenalism is common amongst most of Davidson's critics. Indeed, Kim observes that it has been voiced with "an impressive if unsurprising unanimity" (Kim, 1993a: 20). Moreover, property epiphenomenalism proves to be one of the dividing issues between reductive physicalists, who accept it, and non-reductive physicalists, who generally seek to reject it. The non-reductive physicalist attempts to reject type-epiphenomenalism, and the arguments against them, are the focus of the next two sections.

5. The Non-Reductive Physicalist Defence

The general form of the non-reductive physicalist responses to the charge of type epiphenomenalism (e.g., Macdonald & Macdonald, 1991) is to maintain that we should be careful to distinguish between universals (which they call properties) and particulars (which they call property-instances). A property, e.g. "hardness", is distinct from any given property-instance that instantiates it, e.g. this brick's "being hard", and it is the latter, and not the former, that *are* causally efficacious (it is the brick's "being hard" that breaks the window, not "hardness" in general).

So, whilst mental properties are irreducible to physical ones, particular mental property-instances (that is, mental events which instantiate certain mental properties) may be either realised by (e.g., LePore & Loewer, 1987), or identical[3] to (e.g., Macdonald & Macdonald, 1991), a physical property-instance (that is, the physical event which instantiates certain

universal physical properties).[4] Therefore, whilst mental properties are not causally efficacious, instantiated mental properties are causally efficacious because they are identical to, or realized by, the physical event (which instantiates "its" physical properties).

> 'To say that a mental property of a physical event is causally relevant (that is, that a mental event is causally efficacious *qua* mental) is to say at least that an exemplification of that property, that is, that event, is causally efficacious in bringing about an effect of that event. This will require that (mental) instance to be a physical instance, that is, will require one and the same event to be an instance of both a mental and a physical property' (Macdonald & Macdonald, 1991: 562).

This argument bears an apparent similarity to anomalous monism in so far as it advances a token identity theory. For Davidson, it is not possible to separate mental causation from physical causation precisely because causal relations are between events, and the mental event is identical with the physical event, thus mental causation *just is* physical causation. Similarly, with the property-instance version of non-reductive physicalism, one cannot ask whether a mental event causes a physical event *in virtue of* its instantiating particular mental properties, or in virtue of its instantiating particular physical properties, since the mental property-instance is identical with the physical-property instance, thus mental causation *just is* physical causation.

6. Kim's Overdetermination Critique of Non-Reductive Physicalism

Jaegwon Kim has probably been the most prolific critic of non-reductive physicalism. He charges that non-reductive physicalism, including the property-instance variety outlined above, if it wishes to accept physical causal closure, must either accept type epiphenomenalism or imply causal overdetermination[5] (Kim 1993b, Kim 2005). His argument is as follows:

Suppose that M, a mental event or property-instance, causes another mental event (property-instance), M*. Now, supervenience variously (dependent on its interpretation) implies that both M and M* are identical to *or* realised by physical events (property-instances), let us call these, respectively, P and P* and since M causes M*, P must also cause P*. But, if P causes P*, and M and M* are each, respectively, realised by their physical events, then M* would have been realised irrespective of whether or not it was caused by M, since the instantiation of P* would have realised it. So, says Kim, perhaps the causal efficacy of M comes from its having caused P*. However, if M causes P* then P* is overdetermined; having been caused both by M and by P. Therefore, concludes Kim, either we must accept that mental events cause physical events in virtue of their physical properties (property epiphenomenalism) or we must claim that mental causation always involves the overdetermination of its effects.

Kim originally formulated this argument as one against realization theses such as LePore and Loewer (1987), and, as Macdonald (2007) notes, it does not apply with the same force to the property-instance identity theses.[6] However, where the property-instance identity version fails, Kim would perhaps claim, is in its violation of the explanatory exclusion principle (e.g. Kim, 1989a). The explanatory exclusion principle effectively denies overdetermination in explanation, that is, that there cannot be two independent causal explanations of one event. The property-instance identity thesis fails this criterion because it offers both a mental and physical causal explanation of the same event; thus, Kim might claim, the mental properties cannot be causally relevant (given causal closure of the physical).

Kim summarises the "Mental Causation Problem" for non-reductive physicalists as follows:

> 'Causal efficacy of mental properties is inconsistent with the joint acceptance of the following four claims: (i) physical causal closure, (ii) causal exclusion, (iii) mind-body supervenience, and (iv) mental/physical property dualism—the view that mental properties are irreducible to physical properties' (Kim, 2005: 21-22).

To elaborate briefly on each of these:

i. *Physical Causal Closure* – Any physical effect must have a sufficient

physical cause.

ii. *Causal Exclusion* – "If an event e has a sufficient cause c at t, no event at t distinct from c can be a cause of e (unless this is a genuine case of causal overdetermination[7])" (Kim, 2005: 17).[8]

iii. *Mind-Body Supervenience*[9] – There can be no change in a mental event or property-instance without a change in the corresponding physical event or property-instance.

iv. *Mental/Physical Property Dualism* – Mental properties are irreducible to physical properties (i.e. the distinguishing tenet of non-reductive physicalism).

Kim's position is to reject (iv), maintaining that mental properties are reducible to physical ones, and that mental properties have causal efficacy only because they are so reducible, i.e., mental events are causally efficacious *in virtue of* their physical properties, hence property epiphenomenalism.

7. Anomalous Monism Cannot Imply Property Epiphenomenalism

In his response to Davidson's "Thinking Causes," (in which Davidson defends anomalous monism against its critics), Kim (1993a) argues that simply dispatching with the "inelegant locutions" of "*qua*" and "*in virtue of*"[10] will not get rid of the main issue which, for Kim, '*has always been the causal efficacy of properties of events – no matter how they, the events or the properties, are described*,' (Kim, 1993a: 21 – emphasis in original). This is exactly the same position as is developed by Honderich when he attributes properties to events, and states that the issue of causation is the relation between properties of events. And this is precisely where these critics of anomalous monism, and those who come to its defence with property-based arguments, are wrong. Talk of properties of mental or physical events is just irrelevant to anomalous monism. As Sophie Gibb (2006: 408) so clearly notes, the basic causal relata of Davidson's ontological system are events; properties simply do not feature in it.

7.1 Anomalous Monism, Properties and Predicates

The key to understanding Davidson's position is acknowledging that 'unlike his critics, Davidson does not consider events to have properties, because for him properties are not objective aspects of things in the world' (Gibb, 2006: 414). Davidson endorses a kind of nominalism and rejects a correspondence theory of truth; for him predicates do not pick out objective features of the world that we might call "properties", they do not refer to anything: 'Nothing [...] no *thing*, makes sentences and theories true: not experience, not surface irritations, not the world can make a sentence true' (Davidson, 1974b: 194). Arguably the most important point in understanding the theory of anomalous monism is this denial of the referential character of predicates. So, for Davidson, the statement "this brick is hard" is not true in any sense that involves correspondence with the world.

Indeed, this seems to clearly follow from his argument in support of the Anomalism of the Mental. As noted above, Davidson refers to the 'disparate commitments of the mental and physical schemes' which preclude the possibility of strict psychophysical laws. Now, it is only possible to maintain that both mental and physical descriptions of events can be "true" descriptions of events (as Davidson does) if the criteria of verification, or the *truth-conditions* for the applications of particular mental or physical descriptions, are not inherent in the event itself. For Davidson, properties of events are things we ascribe to them from the perspective of a given theoretical backdrop and vocabulary, 'talk about properties is simply talk about the predicates that can be ascribed to an event when the event is variously described' (Gibb, 2006: 414). We say that the brick *is* hard in relation to its breaking the window only because our physical theory and vocabulary relate those terms in such and such a way, not because, in some objective reality, the brick is hard, or because it instantiates the property of "hardness". Thus, for Davidson, an event could not cause an effect *in virtue of* its having certain properties, since an event need not have (ontologically) *any* properties.

As such, Davidson's denial of type identity is a position of *predicate dualism* – Davidson denies the possibility of conceptual reduction of mental descriptions to physical ones. Ontologically, however, Davidson is an out-and-out monist: there are only events. In contrast, the non-reductive

physicalist maintains, according to Kim (see section 5), a sort of *property dualism*: for non-reductive physicalists mental properties are not *ontologically* reducible to physical ones (although they are somehow dependent on, or realised by, them). Non-reductive physicalists will thus probably be predicate dualists as well, since they take predicates to refer to these properties, but the non-reductivism arises at the ontological, rather than the conceptual level, whilst Davidson argues precisely the opposite; his is a purely conceptual non-reductivism.

7.2 Davidsonian Causal Relations

Davidson's theory of causation maintains that 'causes are individual events, and causal relations hold between events' (Davidson, 1967b: 161). He accepts what he calls the 'principle of extensional substitution' (*Ibid.*: 153) which states simply that causal relations are extensional so that we cannot change the truth-value of a sentence by substituting co-referring terms. In the expression "Alfred's wanting to order caused him to raise his arm", Alfred's wanting to order is a (mental) description of the event that caused the event of Alfred's arm's rising, but another description might be "Alfred's brain state caused him to raise his arm". The latter substitutes a physical description of the first event for a mental one, but, according to Davidson's principle of extensional substitution, since the referent of both expressions is the same event, the truth of the sentence remains unchanged. And this is really the point of Davidson's system: causal relations are between events, *independent* of their description. Since there is nothing in Davidson's ontology for predicates to correspond to, the causal relation cannot be said to be *in virtue of* the event being describable in one way rather than another. As Davidson notes:

> 'Given [my] extensionalist view of causal relations it makes no literal sense [...] to speak of an event causing something as mental, or by virtue of its mental properties' (Davidson, 1993: 13).

7.3 Anomalous Monism Cannot Imply Property Epiphenomenalism

To re-cap, Davidson's ontology does not include properties; his non-reductivism is conceptual rather than ontological. Causal relations are between events and are independent of the manner in which the causal relation is described (or, in strictly non-Davidsonian phraseology, of the properties that the event has). As such Davidson necessarily denies a premise that has been implicit in most of the discussions considered in this paper – that causes have their effects *in virtue of* their properties. For this reason it is not possible to accuse Davidson of rendering mental properties epiphenomenal; properties do no causal work in anomalous monism so it cannot, therefore, be accused of property epiphenomenalism.

Jaegwon Kim, Ted Honderich and the like are well entitled to an ontological system, or "theory of events" (Gibb, 2006: 415) that entails properties, and, indeed, they have advanced such systems themselves (e.g. Kim, 2005). They assume that statements, in some sense, correspond with the world and that physics is the set of truthful statements about reality – in which case mental statements will be truthful only insofar as they are reducible to physical ones. It is fine for them to hold that position, and, indeed, as the argument goes property epiphenomenalism may well be a problem for property dualists of the non-reductive physicalist kind, but Davidson is not a property dualist. It is not fine, however, to criticise anomalous monism because, when placed in this ontological system, it results in property epiphenomenalism, since this is not what Davidson argues – he clearly states his own ontological system that is distinct from the former and from which anomalous monism emerges. As Gibb observes, property epiphenomenalism would be a plausible criticism of 'anomalous monism if embedded within a Kimean theory of events, but to criticise Davidson's theory under a scheme of events that is not his own would be question-begging' (Gibb, 2006: 414-415). As such Davidson and anomalous monism have been erroneously accused of property epiphenomenalism.

Accepting that anomalous monism works with a different ontology means that criticisms of the argument must either be made from within that ontology, or of that ontology itself. Sophie Gibb sets out to attack Davidson's approach to ontology, arguing, instead, that that is really the

problem with anomalous monism. Her arguments are the focus of the next section.

8. On Davidson's Ontology

Sophie Gibb suggests that we should reject anomalous monism *not* because it implies property epiphenomenalism (which, she acknowledges, it doesn't) but because of 'the implausibility of the ontological system within which it is based' (Gibb, 2006: 408). Against Davidson she levels three critiques: that NCC is unsupported, that his nominalism implies the acceptance of disjunctive regularities as strict law statements, and that Davidson does ontology the "wrong way round". This section deals with each of these claims in turn.

8.1 The Nomological Character of Causality is Unsupported

Gibb argues that Davidson does not support his assumption of the Nomological Character of Causality (NCC), which Davidson himself admits. Now, those who accept some kind of correspondence theory of truth, so that strict law statements could correspond to actually existent laws relating cause and effect, might find an ontological justification for NCC, but such a defence is not open to Davidson; who, as we noted above, rejects such correspondence. Indeed the acceptance of NCC really is difficult to reconcile with Davidson's nominalism, but this is a separate line of argument to the claim that his ontology is implausible. Even if it should prove to be the case that the committed Davidsonian cannot maintain NCC, then, of the many other undesirable consequences that could result from this, I do not think the abandonment of the more general stance of anomalous monism is one of them – indeed, we should probably have to abandon physical causal closure, but we need not abandon the claim that the mental interacts causally with the physical, or that there are no strict psychophysical laws. It may be decided that compromising our belief in physical causal closure is too heavy a price to pay in order to accept anomalous monism, in which case let's out with it, but it still remains to be demonstrated that a Davidsonian position actually cannot maintain NCC. Failing to motivate the assumption is a criticism, no doubt, but it only spells serious trouble for anomalous monism, I believe, if it proves to be inconsistent with the nominalism that underpins the theory, and this Gibb does not show.

8.2 Accepting Disjunctive Regularities as Strict Law Statements

Gibb's second argument is that a nominalism of the sort that Davidson embraces would lead to the acceptance of the existence of objective disjunctive regularities. Her discussion starts by asking how the Davidsonian is to identify regularities since, 'without properties, there would seem to be nothing that distinguishes those events that are alike from those that are not' (Gibb, 2006: 418). She suggests that a strategy could be to maintain that events can be distinguished, and regularities consequently identified, by determining the predicates that those events satisfy. The problem with such a position, she suggests, is that it entails that we must accept disjunctive regularities as strict law statements. In this case predicates like "grue" could feature in strict law statements, which, she remarks, is problematic since 'the regularities that this predicate yields, or indeed that any such disjunctive predicate yields, are surely not *real* regularities.' (Gibb, 2006: 419, emphasis added) However, I see no reason why the committed Davidsonian could not bite the bullet here and accept disjunctive regularities of this order as strict law statements: the concern for "realness" only comes in if one accepts correspondence. Indeed, Gibb's concern, as she herself admits, only arises if one accepts a "truthmaker principle" (so that law statements are made true by regularities in the world, i.e. there are "real" regularities which law statements can either correspond or fail to correspond to), Davidson's rejection of such a truthmaker principle effectively shields anomalous monism from the real bite of this argument.

8.3 Davidson's Ontology is "Implausible"

Gibb's final claim is against Davidson's approach to ontology, asserting that 'even given doubts about the truthmaker principle, from an ontological point of view, Davidson has arguably got things the wrong way round.' (Gibb, 2006: 420) She maintains that:

> 'One's motivation for accepting or rejecting an ontological category, and hence a theory of the causal relata, should not have semantic considerations at its base, because contrary to Davidson, a theory of meaning cannot be appealed to in order to settle ontological issues' (*Ibid.*: 420).

Gibb suggests that Davidson ought not to rule out the existence of properties on semantic grounds, and that, instead, metaphysical enquiry should be conducted to establish the nature of causation. This enquiry, she suggests, will find that "properties inevitably play an essential role within one's ontological system and more specifically within one's theory of causation," (*Ibid.*: 420) and, as such, anomalous monism can, and indeed will, be rejected because of the implausibility of its ontological system.

This final argument of Gibb's is the closest to being an actual refutation of anomalous monism – if she were correct and Davidson's ontology were, indeed, misguided, then the arguments for anomalous monism would crumble along with its ontological foundations. However, Gibb provides no actual argument as to why Davidson's ontology is "implausible" or why his approach is the wrong way round, other than repeating the point implicit in the work of reductive and non-reductive physicists alike: that events have their effects in virtue of their properties. Simply saying that 'a theory of causal relata should not have semantic considerations at its base' will not convince the committed Davidsonian, who could as easily reply "Yes, it should."

The difference emerges in the approach of the two camps to metaphysics. As Davidson observes:

> 'When we study terms and sentences directly, not in the light of a comprehensive theory, we must bring metaphysics to language; we assign roles to words and sentences in accord with the categories we independently posit on epistemological or metaphysical grounds. Operating in this way, philosophers ponder such questions as whether there must be entities, perhaps universals, that correspond to predicates, or non-existent entities to correspond to non-denoting names or descriptions' (Davidson, 1977: 205).

This could be said to summarise the key distinction between the kind of approach Gibb advocates (which this quote discusses), and which is perhaps implicit in the property epiphenomenalism debate, and the approach to metaphysics defended by Davidson.

In contrast to Gibb, Davidson takes a holistic approach to the nature of meaning (e.g. Davidson, 1974a). He argues that because we cannot independently separate an agent's beliefs (including our own) from the meaning of the propositions to which those beliefs relate, and because the meaning of any given proposition depends on the system of beliefs into which that proposition is situated, we can only assign meaning to individual propositions once we have something like a theory of interpretation for the language as a whole. For this theory of interpretation Davidson adopts a Tarskian truth-conditional semantics, specifying T-sentences that give truth conditions in the theory language for utterances in the language to be interpreted. The idea is that the total set of T-sentences should maximise agreement between the speaker and the interpreter. In specifying these T-sentences, Davidson notes that we can (and must) give T-sentences which also provide truth conditions for all names and predicates in the language, with the result that we can actually eliminate those semantic terms; 'the call for entities to correspond to predicates disappears when the theory is made to produce T-sentences without excess semantic baggage' (Davidson, 1977: 206). Conversely, Davidson posits the existence of events and people because he maintains that quantifiers must be understood referentially in order to make sense of expressions 'for large stretches of language' (*Ibid.*: 210) (for events, in particular, he argues that this is the only effective way to make sense of adverbial modification (see Davidson, 1967a)). As such Davidson's ontology admits events and agents but does not include properties. This is not to say that he denies their existence, but that he merely argues that they are superfluous to the understanding of ontology, and therefore must also be superfluous to the understanding of causal relations.

Gibb's disagreement is presumably with this approach to ontology, and certainly, it may seem intuitively appealing to posit the existence of entities that correspond to properties, and a long tradition of philosophy has done so. However, Davidson's theory has also been very influential,

and it is insufficient as a criticism of anomalous monism to simply maintain that Davidson's approach to ontology is "the wrong way round", without providing anything in the way of arguments to make that case. A critique of Davidson's semantic approach to metaphysics that proved to be definitive would certainly undermine the cogency of his argument for anomalous monism, but Gibb provides no such critique and, in the end, simply announces her allegiance to the other side.

9. Conclusion

Anomalous monism does not and cannot imply property epiphenomenalism. Physicalists of different stripes are free to battle it out as to whether or not they must accept or reject property epiphenomenalism, but they ought not to involve Davidson in their debates. Anomalous monism is too frequently taken to be a kind of non-reductive physicalism, but in fact the classification runs in the other direction: non-reductive physicalism is a kind of anomalous monism (in so far as there are no psychophysical laws, and it is monistic), but it is a different one to that which Davidson proposes (and, in so far as it accepts property dualism, non-reductive physicalism is arguably a kind of dualism). As Gibb notes, 'it is with good reason that Davidson refers to his position within the philosophy of mind as a monism rather than a physicalism, because for Davidson, events form a neutral class of entities' (Gibb, 2006: 414). Whilst physicalists believe events to be ultimately physical, Davidson makes no such ontological claim; his position is a far more neutral monism than is generally asserted.

There is a lesson to be learned from this debate that Gibb well summarises:

> 'What the problem of mental causation is actually a problem about, and the possible ways of responding to it, depends upon what causation is a relation between; one's theory of the causal relata provides the very framework for one's theory of mental causation' (Gibb, 2006: 407).

In a sense, then, there must be ontological agreement before there can be disagreement about the character of mental causation. Davidson's ontology differs dramatically from those of the physicalists considered in this paper; as such anomalous monism cannot answer to the problems that affect those systems. More refined criticisms must face against the ontology directly, or treat anomalous monism on its own terms.

Deren Olgun (1985) completed the EIPE Research Masters in Philosophy and Economics in August 2011. His research interests are in the intersections of the philosophy of social science, philosophy of action, philosophy of mind, metaphysics and the philosophical foundations of decision theory.

'The Argument for Anomalous Monism, Again' was written for the mastercourse 'Experimenting Ethics Away' (of dr. Maureen Sie), taught by dr. Leon de Bruin.

Literature

Crane, T. (1995) 'The Mental Causation Debate' In: *Proceedings of the Aristotelian Society*, 69: 211–236

Crane, T. (2003) 'Mental substances'. In A. O'Hear (ed.), *Minds and persons*. Cambridge: Cambridge University Press: 229-250

Davidson, D. (1963/2001) 'Actions, Reasons, and Causes'. In: *Essays on Actions and Events*. New York: Oxford University Press: 3-21.

Davidson, D. (1967a/2001) 'The Logical Form of Action Sentences'. In: *Essays on Actions and Events*. New York: Oxford University Press: 105-122.

Davidson, D. (1967b/2001) 'Causal Relations'. In: *Essays on Actions and Events*. New York: Oxford University Press: 149-162.

Davidson, D. (1969/2001) 'Individuation of Events'. In: *Essays on Actions and Events. New York*: Oxford University Press: 163-180.

Davidson, D. (1970/2001) 'Mental Events'. In: *Essays on Actions and Events*. New York: Oxford University Press: 207-224.

Davidson, D. (1974a/2001) 'Belief and the Basis of Meaning'. In: *Inquiries into Truth and Interpretation*. New York: Oxford University Press: 141-154.

Davidson, D. (1974b/2001) 'On the Very Idea of a Conceptual Scheme'. In: *Inquiries into Truth and Interpretation*. New York: Oxford University Press: 183-198.

Davidson, D. (1977/2001) 'The Method of Truth in Metaphysics'. In: *Inquiries into Truth and Interpretation*. New York: Oxford University Press: 199-214.

Davidson, D. (1993). 'Thinking Causes'. In J. Heil, & A. Mele (Eds.), *Mental causation*. Oxford: Clarendon Press.

Gibb, S. (2006) 'Why Davidson is Not a Property Epiphenomenalist'. In: *International Journal of Philosophical Studies*, 14(3): 407-422.

Honderich, T. (1982) 'The Argument for Anomalous Monism'. In: *Analysis*, 42(1): 59.

Jacob, P. (2002) 'Some Problems for Recductive Physicalism'. In: *Philosophy and Phenomenological Research*, 65: 648–654

Kim, J. (1989a) 'Mechanism, Purpose, and Explanatory Exclusion'. In: *Philosophical Perspectives*, 3: 77-108.

Kim, J. (1989b) 'The Myth of Non-reductive Materialism'. In: *Proceedings and Addresses of the American Philosophical Association*, 63(3): 31-47.

Kim, J. (1993a) 'Can Supervenience and Non-strict Laws Save Anomalous Monism?'. In J. Heil, & A. Mele (Eds.), In: *Mental causation.* Oxford: Clarendon Press: 19-26

Kim, J. (1993b) 'The Non-reductivists' Troubles with Mental Causation'. In J. Heil, & A. Mele (Eds.), In: *Mental causation.* Oxford: Clarendon Press: 189-210

Kim, J. (1996) *Philosophy of Mind*. Oxford: Westview Press.

Kim, J. (2005) *Physicalism, or Something Near Enough*. Princeton: Princeton University Press.

Le Pore, E., & Loewer, B. (1987) 'Mind Matters'. In: *The Journal of Philosophy*, 84(11): 630-642.

MacDonald, C., & MacDonald, G. (1986) 'Mental causes and Explanation of Action'. In: *The Philosophical Quarterly*, 36(143): 145-158.

Macdonald, G. (2007) 'Emergence and Causal Powers'. In: *Erkenntnis*, 67(2): 239-253.

McLaughlin, B. (1993) 'On Davidson's Response to the Charge of Epiphenomalism'. In J. Heil, & A. Mele (Eds.), *Mental causation*. Oxford: Clarendon Press: 27-40

Notes

1 A position which is implied by NCC in conjunction with AOM.

2 As, Honderich suggests, is implied by NCC.

3 These notions of realisation and identity are associated with the supervenience thesis, originally entailed in anomalous monism, specifically that the mental supervenes on the physical.

4 "Exemplifications of mental properties of mental events are identical with exemplifications of physical properties of physical events" (Macdonald and Macdonald, 1991: 562).

5 Overdetermination, simply put, is where a given event is caused by two or more independent causes that would each, by themselves, be sufficient to bring about the event. Consider, for example, a (non-waterproof) watch that is simultaneously immersed in water and crushed in a vice. Both the immersion in water and the crushing of the watch would be sufficient for the watch to seize functioning. The event of the watch ceasing to function is therefore said to be overdetermined. It does seem counter-intuitive to suppose

that overdetermination plays such a continuous role in our experience as to be occurring whenever the mental interacts causally with the physical, and, indeed, Kim calls such a position "absurd" (Kim, 1989b: 44).

6 Because the identity argument maintains that mental causation is physical causation, the two causes are not independent but are the same, therefore there is no overdetermination.

7 Consider the watch example above as a "genuine" case.

8 It may be worth distinguishing here between causal exclusion and explanatory exclusion. Whilst causal exclusion is an ontological thesis, claiming that no one event can have two independent causes, explanatory exclusion is more of an epistemological thesis, saying that no one event can have two independent explanations. In effect the former argues that a mental event and a physical event cannot both cause a single physical event, whilst the latter argues that the same event cannot be given both a mental and physical explanation. To make the case highlighted here Kim only uses causal exclusion.

9 The formulation of supervenience is a particular thorny issue in the context of this debate, but for the purposes of this paper nothing hangs on the nature of supervenience, therefore this definition will suffice.

10 Which is Kim's understanding of Davidson's intention in *Thinking Causes*.

ESJP
#1 | 2011

Geschreven met licht – *The Falling Man* en de vervluchtiging van de tragedie in de geest van de rouw

Karin de Bruijn

The force behind the movement of time
is a mourning that will not be comforted.
That is why the first event is known to have been an expulsion,
and the last is hoped to be a reconciliation and return.

– Marilynne Robinson

De laatste momenten in het leven van een mens spreken tot de verbeelding, zeker wanneer deze zich ontvouwen in het brandpunt van de actualiteit. Denk bijvoorbeeld aan de foto die op het internet werd *getagd* als *The Falling Man*. Het beeld toont, in de meedogenloze eerlijkheid van een cameralens, iets wat op de dag zelf door mensenogen niet waargenomen had kunnen worden. *The Falling Man* is één van de *jumpers* van *9/11*, de mensen die de huiveringwekkende val van een hoogte van meer dan honderd verdiepingen verkozen boven de hel van de brandende Twin Towers. Als de wereld al stilstond op 11 september 2001, dan was het zeker in de beklemmende stasis van dit beeld. Losgezongen van de achtergrond van de catastrofale gebeurtenissen en van zijn eigen wanhoopsdaad, lijkt *The Falling Man* gevangen tussen leven en dood, vereeuwigd in de fractie van een seconde waarin zijn vrije val zich voltrekt. Wégkijken – een instinctieve reactie – heeft geen zin. Ook de kijker is gevangen in het beeld dat na één enkele aanschouwing voor altijd op het netvlies staat gebrand. En bij iedere 'heraanschouwing' is de kijker genoodzaakt stil te staan bij de wanhoop en de eenzaamheid van de laatste momenten van deze mens.

Aan de vooravond van de tiende herdenkingsdag van *9/11* domineerden de laatste momenten van een ander mens voor even de internationale media. De beelden van die momenten werden weliswaar niet vrijgegeven, maar dat laat onverlet dat er met de executie van Osama Bin Laden alweer een zwarte bladzijde van de geschiedenis werd omgeslagen. De beelden in de media van juichende Amerikanen riepen herinneringen op aan de beelden van juichende mensen in andere delen van de wereld daags na *9/11*.

Zonder al te grote overdrijving kunnen we stellen dat de eenentwintigste eeuw is begonnen op 11 september 2001, maar tegelijkertijd kunnen de tragische gebeurtenissen van die dag worden opgevat als niet meer dan één enkele manifestatie van de meer omvattende crises – sociaal, ecologisch, economisch – waarmee de mensheid in de geglobaliseerde samenleving van nu te maken heeft. Welk licht laten we schijnen op het eerste decennium van deze nieuwe eeuw, en op welke wijze laten wij ons aanspreken door de voornoemde beelden die daar het begin- en eindpunt van markeren? De triomf van de overwinnaars en de rouw van de verliezers, en de wijze waarop winst en verlies, rouw en triomf tussen actoren verschuiven, doen steeds een andere dimensie van de gebeurtenissen oplichten. Welke schaduwen worden hierdoor vooruitgeworpen naar de toekomst?

In haar analyse van de gebeurtenissen en de nasleep van *9/11* vraagt Judith Butler onze aandacht voor de politieke dimensie van de rouw. Om het verlies van sommige levens kan en mag publiekelijk gerouwd worden, om het verlies van andere levens niet, aldus Butler. Dit alom aanwezige maar nauwelijks geproblematiseerde besef heeft politieke consequenties. De rouw stelt ons, zo schrijft Butler, open voor de kwetsbaarheid van onszelf en van anderen, en voor het lijden dat altijd een reële mogelijkheid is die met die kwetsbaarheid gegeven is. De ander kan één van ons zijn, zoals de mens in het beeld van *The Falling Man*, maar de ander kan zich ook elders in de wereld bevinden, zoals de 'naamloze' slachtoffers van

een politiek die ver weg maar mede uit onze naam wordt gevoerd. De geschiedenis van de mensheid laat zich immers vertellen als een grillig en meerstemmig verhaal – een verhaal van strijd, van winnaars en verliezers, van 'goed' en 'kwaad', van vooruitgang en neergang – allemaal meer of minder relatieve kwalificaties. De toeschrijving van die kwalificaties aan specifieke gebeurtenissen en actoren is sterk ingekleurd door het perspectief van diegene die het verhaal vertelt. Wat echter nooit gerelativeerd kan en mag worden is de onderstroom van die geschiedenis: het lijden van mensen van vlees en bloed, en de rouw die hiermee gepaard gaat. Dit besef zou ons er volgens Butler bovendien van moeten weerhouden om op de uitschakeling van degene die wij als de vijand beschouwen met uitingen van triomf en genoegdoening te reageren. De rouw verbindt mensen met elkaar, over tijd, plaats en politieke en morele conflicten heen, en stelt hen, iedere keer opnieuw, voor een politieke keuze: de keuze voor een strategie van strijd en vergelding, of de keuze voor een strategie van matiging en verzoening.

Butler grijpt in haar werk terug op *Antigone*, een tragedie van de hand van de Griekse tragedieschrijver Sophocles. Zij sluit daarmee aan bij een rijke traditie in de westerse filosofie en kunsten die de *condition humaine* in de eigen tijd tracht te duiden door zich tot één van de bakermatten van de westerse cultuur te wenden: het Athene van ruim 2.500 jaar geleden. Hier bereikt de esthetisering van het lijden van de mens een eerste hoogtepunt in het kroonstuk van het Griekse erfgoed: de tragedie. Eén van de beroemdste voorbeelden van de wijze waarop de tragedie filosofen en kunstenaars over de eeuwen heen aanspreekt, is de interpretatie van Friedrich Nietzsche. Nietzsche hoopte op een revitalisering van de Europese cultuur door de wedergeboorte van de tragedie uit de geest van de muziek. In onze tijd verdedigt Jos de Mul de stelling dat de wedergeboorte van het tragische zich inderdaad heeft voltrokken, maar niet in het domein van de kunsten, zoals Nietzsche viseerde, maar uit de geest van de technologie.

Het Athene van de grote tragedieschrijvers is een hogedrukpan gevuld met ingrediënten die bepalend zullen worden voor de ontwikkeling van de westerse cultuur: het rationele denken, de wetenschap, de filosofie. Maar ook: oorlog, kolonisatie, slavernij. Een bruisend laboratorium, op het breukvlak tussen twee werelden: de wereld van de *mythos*, waarin de hoofdrollen nog zijn weggelegd voor onbuigzame natuurkrachten en wispelturige goden waaraan zelfs de meest heroïsche stervelingen zijn overgeleverd, en de wereld van de *logos*, waarin de nieuw ontdekte waardigheid en vrijheid van de mens op de proef worden gesteld in het politieke experiment van de democratie. Balancerend op dit breukvlak staat de tragische held. Deze held ziet zich gedwongen tot het maken van een onmogelijke, en daarmee noodlottige, keuze. De keuze is onmogelijk, omdat het een keuze is tussen twee imperatieven die voortkomen uit verschillende morele ordes die aan de mens voorafgaan en die hem overstijgen, en die daarom een niet te verloochenen aanspraak op hem maken. De imperatieven zijn tegengesteld en onverzoenlijk. Daarmee is de keuze die de held moet maken per definitie noodlottig. Welke keuze hij ook maakt, de held zal één van de imperatieven, en daarmee één van de morele ordes, moeten verloochenen, en hij zal daarvoor de volle verantwoordelijkheid dragen. Het is deze keuze'vrijheid', waarin vrijheid en noodzakelijkheid samenvallen, die de held tot een tragische held maakt. In de ontwikkeling van de Griekse tragedie voltrekt zich daarbij een subtiele ontwikkeling waarbij het karakter van de held een steeds grotere rol in het keuzeproces gaat spelen. De meest tragische van de tragische helden is degene die gedreven wordt door een *daimon*, een kracht die op hem inwerkt en die hij zich willend toeëigent.

Creon en Antigone, de tragische tegenspelers uit de tragedie van Sophocles, zien zich geconfronteerd met een conflict tussen de imperatief van de *polis*, de politieke gemeenschap die de mensen in de publieke ruimte met elkaar verbindt, en de *oikos*, de gemeenschap tussen mensen die wortelt in verwantschap en liefde. In *Antigone* wordt de keuze van de helden op de spits gedreven door hun halsstarrigheid. Creon verbiedt zijn onderdanen om publiekelijk te rouwen voor Polyneikos, die gesneuveld is in de burgeroorlog tegen zijn broer Eteokles. Het decreet wordt bekrachtigd door Polyneikos postuum tot vijand te bestempelen en hem ex-polis te verklaren; zijn stoffelijk overschot wordt buiten de stad geplaatst en aan de aaseters overgelaten. Antigone, de zus van Polyneikos en Eteokles, legt het decreet van Creon naast zich neer en bewijst haar broer symbolisch de laatste eer door een handvol aarde over zijn lichaam uit te strooien. Creons onvermurwbaarheid – hij weigert te luisteren naar degenen die hem smeken om zijn decreet te versoepelen - vindt een evenknie in Antigone's

halsstarrigheid; zij bewijst Polyneikos niet één maar twee maal de laatste eer, de tweede keer in het zicht van haar stadgenoten. Het resultaat is voor Creon en Antigone en voor alle andere tragische helden gelijk; in het streven naar het goede stevent de held onafwendbaar op zijn zelfgekozen noodlot af en trekt daarbij in de meeste gevallen een spoor van lijden en rouw door zijn eigen gemeenschap.

In hoeverre kan de tragische levensvisie zoals gearticuleerd in *Antigone* en andere Griekse tragedies nog een licht werpen op de *condition humaine* in de eenentwintigste eeuw, en op de even iconische als anecdotische wijze waarop deze zich manifesteert in de gebeurtenissen van *9/11* en in het noodlottige einde van *The Falling Man*? Het decor van de Griekse polis is door de compressie van tijd en afstand getransformeerd in een *global village* waarin zeven miljard mensen met elkaar samenleven. Het primaat van het tragische ligt niet langer bij de enkeling, maar bij anonieme processen die een ontkenning inhouden van de morele ordes waarin de Griekse mens zichzelf geplaatst zag en die uitsluitend onderhavig lijken te zijn aan hun eigen immanente logica, de logica van het kapitaal en de technologische vooruitgang. Ingevoegd in dit proces is de mens tegelijkertijd vrij en onvrij. Schuld en verantwoordelijkheid, macht en onmacht, zijn collectief geworden, onevenredig verdeeld over mensen en niet langer exclusief en ondubbelzinnig aan specifieke actoren toe te schrijven. Het tragische is systemisch en - gezien de ecologische, sociale en economische crises van onze tijd - pathologisch geworden. Net zoals de helden uit de Griekse tragedies dragen de processen de kiem van hun eigen vernietiging in zich. De mens in de *global village* lijkt daardoor onafwendbaar op een toekomst af te stevenen die wel eens een noodlottig einde, of tenminste een radicale breuk zou kunnen inhouden.

De werkelijkheid waarin de mens van de eenentwintigste eeuw zich geplaatst ziet is weliswaar radicaal verschillend van de wereld van de oude Grieken, net zoals er grote verschillen bestaan tussen hun en ons zelfbeeld, maar dat wil niet zeggen dat de tragedieschrijvers ons niets meer te zeggen hebben; al was het maar omdat zij ons een spiegel voorhouden waarin we een begin van de contouren van ons moderne zelfbeeld kunnen ontwaren. Echter, door onze focus op de tragische held wenden wij wellicht onze blik teveel af van datgene wat zich rondom de held afspeelt. Wij zijn in de ban van de tragische held en diens strijd tegen de grenzen en beperkingen die hem door de wetten van de goden, de natuur of de sociale gemeenschap worden opgelegd. Strijdbaarheid in het licht van dergelijke beperkingen is uiteraard een belangrijke voorwaarde voor de emancipatoire processen waarvan in de recente geschiedenis veel mensen hebben geprofiteerd, maar de strijd in de Griekse tragedies is een strijd die vaak ten koste gaat van iets of iemand. In veel gevallen ten koste van de held zelf, maar vaak ook ten koste van anderen. Het lijden dat in de tragedie wordt geësthetiseerd, is vooral de innerlijke strijd van de tragische held. In lijn met de heldenethiek van de Grieken worden het fysieke lijden en de dood van de held veeleer als bijzaak gepresenteerd. De tragedie lijkt bovendien weinig oog te hebben voor het fysieke lijden dat zich achter de coulissen en buiten het gezichtsveld van het publiek afspeelt. De medespelers worden gedood of slaan, door rouw en wanhoop gedreven, de hand aan zichzelf, en vormen zo de *collateral damage* van het handelen van de hoofdrolspelers.

Die veronachtzaming van de *collateral damage* lijkt een belangrijk punt van verbinding tussen onze tijd en die van de oude Grieken. Of het nu, in onze tijd, gaat om het competitieve individu in het speelveld van de (neo)liberale economie of om de strijder die zich met geweld inzet voor de 'goede zaak', beiden kijken weg van de destructie die in de *slipstream* van hun handelen ontstaat – de destructie van het leven van mensen en van het leven op aarde waarin de mens is ingebed. Maar er is ook een belangrijk verschil. In zijn optimistische vooruitgangsgeloof heeft de moderne mens het tragische besef van de eigen kwetsbaarheid en sterfelijkheid verdrongen en vervangen door het optimistische geloof in de vooruitgang en de beloftes van de techniek. De *hybris* van de moderne mens gaat alle grenzen te buiten. De moderne mens, ingebed in processen waar hij de grip op verloren heeft, lijkt door 'doemsvergetelheid' gedreven de vrijheid te ruim en te absoluut te hebben opgevat en de zwaartekracht van de noodlottigheid te hebben miskend.

Echter, de verheerlijking van de tragische held en de veronachtzaming van het lijden van de minder heroïsche medemens werden door de Grieken wellicht veel sterker gerelativeerd en geproblematiseerd dan in de moderne tijd. Voor de Grieken was de tragedie slechts één onderdeel van een veel rijker geschakeerd wereldbeeld en van een rijkere constellatie van

praktijken. De tragedies werden geschreven en opgevoerd als een trilogie van drie met elkaar verwante stukken. De enige compleet overgeleverde trilogie, de *Oresteia* van Aeschylus, kent in tegenstelling tot veel afzonderlijke tragedies een relatief gelukkig einde. De opvoering van de tragedies werd afgesloten met de opvoering van een komedie, waarin veel van de uitgangspunten van de tragedies, en vaak ook de tragedieschrijvers zelf, behoorlijk op de korrel werden genomen. Ook de grote verschillen tussen de tragische helden en het rijke pantheon van goden waaruit de Grieken konden putten, plaatsen de tragedies in een ander licht, net als de opvattingen van Aristoteles, die de tragedies toch vooral beschouwde als fraai vormgegeven levenslessen die de mens op het rechte pad moesten houden.

In onze interpretatie van de tragische levensvisie van de Grieken zouden we bovendien veel meer de nadruk kunnen leggen op de rol van de noodlottigheid, waaraan de rouw ontspringt, en op de kracht van de verzoening, waarin de rouw mogelijk uitmondt. In *Antigone* komen de existentiële en politieke dimensies van de rouw samen. De rouw in zijn existentiële dimensie komt impliciet aan de orde in de eerste koorzang, waarin we, in de vertaling van Martin Heidegger, het koor horen zeggen dat er vele zaken zijn die *unheimlich* zijn, maar dat er niets zo *unheimlich* is als de mens. *Unheimlich* betekent onheilspellend of sinister en zou in een vrije interpretatie opgevat kunnen worden als 'verdoemd'. De mens is het verdoemde, het noodlottige dier, en het dier dat zich van de eigen noodlottigheid bewust is. In deze noodlottigheid vallen niet alleen de absolute contingentie van zijn eigen geboorte en de absolute noodzaak van zijn eigen dood samen, maar ook de mogelijkheid om zowel het goede als het kwade te doen – de mens is daarmee ook het vervloekte dier. *Unheimlich* kan volgens Heidegger ook worden opgevat als 'ontheemd' of 'thuisloos'; de mens is als sterfelijk wezen niet thuis in het Zijn. In het verlengde hiervan kunnen we ons de vraag stellen of deze existentiële ontheemdheid in de mens ook niet de heimwee oproept, het verlangen naar een thuis, naar een *oikos*, een leefgemeenschap waarin hij in verbinding en harmonie met anderen beschutting vindt. De medespelers van Creon en Antigone pleiten voor de strategie van matiging en verzoening die voor deze harmonie van levensbelang is. In de context van *9/11* stelt de rouw in zijn existentiële dimensie ons open voor het belang van de heelheid van ons thuis en van het thuis van anderen. De vernietiging van een deel van de *oikos* die New York heet, dwingt ons stil te staan bij de vernietiging, óók uit onze naam, van het thuis van anderen – niet alleen van andere mensen, maar ook van andere levende wezens. De rouw in zijn existentiële dimensie stelt ons, met Butler gesproken, open voor de rouw in zijn politieke dimensie. Zo eist Antigone het recht op om publiekelijk te rouwen om Polyneikos. Elk leven is het rouwen waard. De rouwers dienen bovendien in hun rouw gerespecteerd te worden, ook als zij rouwen om iemand die wij als onze vijand beschouwen. Natuurlijk kan in dat geval niet van ons worden verwacht dat wij delen in die rouw, maar het past ons evenmin om ons uit te leven in uitingen van triomf en genoegdoening.

Zo gelezen, bieden de Griekse tragedies en de bredere context waarin zij begrepen kunnen worden ons een alternatief perspectief op het leven in de eenentwintigste eeuw. Dit is een perspectief waarin de verabsolutering van de menselijke vrijheid en de verheerlijking van de heroïsche strijd op zijn minst worden getemperd door de erkenning dat ook en juist in onze tijd de onvrijheid en de noodlottigheid van de mens schreeuwen om strategieën van matiging en verzoening. In het verlengde van de these van De Mul dat in onze tijd het tragische is wedergeboren uit de geest van de techniek, kunnen we stellen dat we een nieuw perspectief nodig hebben. Een perspectief waarin de tragedie als het ware oplost in de rouw, waarbij een essentieel onderdeel van de tragische levensvisie van de oude Grieken resteert: het besef van de eigen kwetsbaarheid en sterfelijkheid.

Net als Judith Butler in de context van de aanslagen van *9/11*, wijst Slavoj Žižek in de context van de ecologische, sociale en economische crises van de eenentwintigste eeuw op de cruciale rol van het rouwproces voor de levensvatbaarheid van onze toekomst. Met *The Falling Man* doen zij een beroep op ons om níet weg te kijken, niet van onze eigen kwetsbaarheid en evenmin van het lijden dat in de *slipstream* van ons handelen ontstaat. Want ofschoon we in de virtuele wereld van massamedia en internet in *realtime* worden geconfronteerd met het lijden van mensen en andere levende wezens en met de vervuiling en vernietiging van ons en hun habitat, toch moeten we onszelf de vraag stellen of we werkelijk kunnen en willen zien – of kunnen en willen voelen – wat dit betekent. Ontkenning, marchanderen, woede, en depressie, de fasen in het rouwproces die door Žižek worden aangewend in zijn politieke analyse, dienen in de rouwverwerking

te worden doorlopen voordat ze kunnen worden omgezet in acceptatie, als eerste stap op weg naar duurzame verandering. Wat de mens in de eenentwintigste eeuw nodig heeft is niet een esthetisering van het lijden, zoals in de Griekse tragedie, maar een esthetisering van de rouw – een esthetisering die ons helpt om de rouw in zijn existentiële en politieke dimensies te doorleven en ons open te stellen voor strategieën van matiging en verzoening. Een dergelijke esthetisering is zichtbaar in de kunst en de literatuur die in de nasleep van *9/11* tot stand is gekomen, en is bijvoorbeeld ook nodig met het oog op de ecologische crisis. Zo vraagt Ilija Trojanow zich af hoe hij een roman kan schrijven over de klimaatcrisis, en hij laat zich in zijn vraag leiden door zijn droom over een gletsjerexpert die rouwt om de teloorgang van zijn geliefde gletsjer. Als zwart de kleur is van de rouw, dan wordt de esthetisering van de rouw, door de kunst en de filosofie, daarin ingeschreven met licht.

Karin de Bruijn (1964) studeert Wijsbegeerte aan de Erasmus Universiteit Rotterdam. In 1997 behaalde zij haar doctoraaldiploma (cum laude) in de Engelse taal- en letterkunde aan de Rijksuniversiteit Leiden (scriptie: Silent Conversations: The Reconstruction of (Human) Nature in Marilynne Robinson's Housekeeping and Gloria Naylor's Mama Day). Onderzoeksinteresses: filosofie van mens en cultuur, ecologische filosofie, critical animal studies.

'Geschreven met Licht' is geschreven ter afronding van het bachelorvak 'De domesticatie van het noodlot' van prof. dr. Jos de Mul.

Literatuur

Butler, Judith (2004) *Precarious Life: The powers of mourning and violence*. Londen/New York: Verso.

Kroes, Rob (2009) 'Vallende Man: Een blijvend beeld van 9/11.' In: *De Groene Amsterdammer*, 133, 35, 36-39.

Mul, Jos de (2006) *De domesticatie van het noodlot: De wedergeboorte van de tragedie uit de geest van de technologie*. Kampen: Klement/Pelckmans.

Sophocles (1891) *The Antigone of Sophocles*. (Sir Richard Jebb, Ed.). Cambridge: Cambridge University Press. Te raadplegen via www.perseus.tufts.edu.

Trojanow, Ilija (2010) *Requiem voor de toekomst: Hoe schrijft men een roman over de klimaatramp*. Amsterdam: Volkskrant Boekenfonds.

Žižek, Slavoj (2010) *Living in the end times*. Londen/New York: Verso.

ESJP
#1 | 2011

Ideal Theory and Utopia

Volker Ruitinga

Introduction

Social and political philosophy has inherited a great deal from the work of John Rawls. He has been widely credited with reviving social theory in the 20th century. Amongst his many contributions is an intuitively simple concept which has sparked controversy and debate in the post-Rawlsian literature on theories of justice: that of Ideal Theory, as opposed to non-Ideal Theory. What seems like a simple distinction has become the subject of debate for two important reasons: firstly, Rawls is not particularly concerned with the distinction and leaves it ill-defined. To illustrate: in his work *A Theory of Justice* the words 'ideal' and 'theory' appear in conjunction a mere five times. Yet, the Ideal / non-Ideal distinction strikes at the heart of social theory. If we take the basic goal of social theory to be advancing the betterment of society, many scholars agree that some ideal conception of society is of value to this end (see Bloch, 1986; Davis, 1981; Levitas, 2011). Secondly, even if Rawls had provided a clear definition of the distinction between Ideal and non-Ideal Theory, controversy surrounding the object of the distinction would remain.

Rawls defined his own work as Ideal Theory. Consequently, the literature on social and political philosophy has seen several attempts to categorize theories of justice based on the Ideal/non-Ideal distinction (e.g. Stemplowska, 2008); an obvious problem with this endeavour being that there is no consensus on what the definition of this distinction may be. Therefore, many of these efforts have only succeeded in highlighting certain difficulties surrounding the debate. This goes to show that both the concept of Ideal Theory and therewith the supposedly 'ideal theories' themselves are the source of much disagreement. Thus, we must examine the nature of Ideal Theory, its contours and defining characteristics, before we can constructively discuss the consequences of this distinction on the theories in question.

I wish to start by suggesting that all engagement in ideal theorizing has at its source some conception of an ideal society. Outside the post-Rawlsian literature on theories of justice this concept has often been referred to as *Utopia*. However, in the literature on Rawlsian Ideal Theory there is a marked absence of utopian terminology. One of the reasons for this absence may be that scholars are wary of using the term because it is itself the subject of much controversy. Another reason may be that the term is traditionally used to refer to more than issues of justice alone. However, although these are defensible grounds for introducing new terminology, unique to theories of justice, ignoring utopian literature results in a number of missed opportunities. Firstly, there is undoubtedly significant overlap between Utopian Theory and Ideal Theory. Ideal theories of social justice can potentially be categorized as a subset of Utopian Theory, alongside for example Utopian- Moral, Political and Legal Theory. Secondly, owing to this overlap, examining the literature on Utopian Theory may reveal significant parallels between the two debates. Thirdly, and most importantly, integrating the concept of utopia into the definition of Ideal Theory could serve to clarify the concept. This will, in turn, lead to a more fruitful use of the distinction.

In light of this observation the aim of this paper is twofold. Firstly, to suggest a new way of understanding Ideal Theory, specifically with reference to utopian literature, in order to illustrate the benefit of introducing the concept Utopia to the existing literature on Ideal Theory. To this end I will re-examine the Rawlsian definition of Ideal Theory and briefly sketch the similarities between Ideal Theory and Utopia. I aim to show that Ideal

Theories of Justice should be regarded as Utopian. Secondly, I will look at two articles that discuss the relevance of Ideal Theory to see if my proposed amendments aid the defence of Ideal Theory against criticism, as I consider Ideal Theory and Utopia to be essential to social theory.[1]

Part I

In order to compare utopianism with Rawlsian Ideal Theory, a brief introduction to Utopia is required. In 1516 Sir Thomas More gave the name Utopia to his proposed ideal society. At that time the word was a neologism, or more accurately a lexical neologism, meaning that it named a new concept or synthesized pre-existing ones (Vieira, 2010: 3). It combined the Greek words ou (οὐ), meaning *no*, and topos (τόπος), meaning *place*, adding the suffice '*ia*' indicating *a* place. More originally intended to use the word Nusquama, nusquam being Latin for 'nowhere' or 'never'. However, he chose Utopia, for when spoken in English both Outopia and Eutopia are phonetically alike, the latter meaning *good* place. This eloquently captured the ambiguity of any such imagined society, the unceasing tension between the ideal and the unreachable.

Utopian *Theory* is essentially concerned with conceptualizing the 'ideal commonwealth', which inherently expresses a 'psychological aspiration of hope for a better state of existence in this life or elsewhere, notably in the form of the quest for "community"' (Claeys, 2010: xi). Often this 'theorising' has taken the form of literature, as in Thomas More's *Utopia*. But Utopian Theory encompasses much more than enticing storytelling alone. Karl Mannheim famously wrote on the relationship between Utopia and ideology (Mannheim, 1954), Bloch on Utopia and hope (Bloch, 1985), Goodwin on Utopia and politics (Goodwin, 2009), and so on. Utopian Theory encompasses many aspects of social theory, because Utopia is so fundamental to our thinking about society. With this in mind, we turn to the examination of Rawlsian Ideal Theory.

Rawlsian Ideal Theory and Utopia

There are many interpretations of Ideal Theory. For example, some scholars (erroneously) equate Ideal Theory solely with the condition of full-compliance (to the conditions prescribed for society), one problem with this equation being that full compliance 'may also hold for principles of justice which do not lead to a just society' (Robeyns, 2008: 3). Moreover, nowhere does Rawls say that it is a sufficient condition for Ideal Theory. However, this does beg the question: what conditions are necessary for an Ideal Theory according to Rawls? In order to answer this we must highlight a few key passages regarding Ideal Theory from *A Theory of Justice*.

To start, Rawls limits the scope of his inquiries in several respects. For instance, he is concerned only with instances of justice, for 'justice is the first virtue of social institutions' (Rawls, 1999: 3). A further limitation is best summarized by the following: 'I shall be satisfied if it is possible to formulate a reasonable conception of justice for the basic structure of society conceived for the time being as a closed system isolated from other societies' (Rawls, 1999: 7). Here Rawls emphasises the need for theoretical simplification by stating that his ideal society is both closed and isolated. Rawls also writes: '[...] I consider primarily what I call strict compliance as opposed to partial compliance theory' (Rawls, 1999: 7). Crucially, he goes on to say that *partial* compliance theory 'studies the principles that govern how we deal with injustice' (Rawls, 1999: 7), meaning that when there is full-compliance to the hypothesized principles of the theory of justice there can be no instances of injustice.

So far we know that Ideal Theory is concerned with providing a theory of justice for an isolated society, whose members act in full compliance to the proposed principles of justice, resulting in a situation where there are no instances of injustice. This leaves a very narrow definition of Ideal Theory, as any theory concerning partial compliance or indeed one that results in any instances of injustice (resulting from proposed institutional policy) would not be considered an Ideal Theory. Rawls goes on to defend this view by stating: 'The reason for beginning with ideal theory is that it provides, I believe, the only basis for the systematic grasp of these more pressing problems. […] At least, I shall assume that a deeper understanding can be gained in no other way, and that the nature and aims of a perfectly

just society is the fundamental part of the theory of justice' (Rawls, 1999: 8). Hence, Rawls claims that there can be no complete non-Ideal Theory without Ideal Theory.

Rawls comes closest to a definition of Ideal Theory in the following passage:

> 'The intuitive idea is to split the theory of justice into two parts. The first or ideal part assumes strict compliance and works out the principles that characterize a well-ordered society under favorable circumstances. It develops the conception of a perfectly just basic structure and the corresponding duties and obligations of persons under the fixed constraints of human life' (Rawls, 1999: 216).

Here, Rawls points out that the endeavour of designing an Ideal Theory requires certain theoretical limitations. One of which is that it requires the paradoxical assumption of favourable circumstances whilst accepting the fixed constraints of human life; (paradoxical because the constraints of human life are often unfavourable to creating an ideal society).

At this point it must be noted that many of these same conditions and limitations also hold for conceptions of Utopia. Utopian Theory often envisions an isolated society, under favourable conditions, where there can be no injustice if its members comply to the societal ideals. This goes to show that the definition of Rawlsian Ideal Theory arguably holds as a viable, although simplified, definition of Utopia. We can then define Ideal Theory to be: *a system of principles that, when fully-complied to by all members of society, results in a Utopia of social justice*.

Ideal Theory and Utopianism

Against this backdrop, a number of significant similarities between Utopian Theory and Ideal Theory become apparent. For instance, both Utopian Theory and Ideal Theory envision some significantly improved version of society. Furthermore, neither theory is primarily concerned with explicating the transition to this ideal from our world.

But there are other, more subtle parallels, for example: both theories attach some specific value to their imagined place. In Eutopia this is the *good*, in Ecotopia, Vegatopia and Technotopia their overriding values are clear. In this vein, Rawslian Ideal Theory is primarily concerned with *justice*. Additionally, they share the same potential to guide human progress by presenting a well-argued example. Conversely, both Utopian Theory and Ideal Theory are subject to many of the same criticisms: their ideas are said to be unreachable fantasies and to pursue them is a waste of time. Furthermore, it is no coincidence that Sir Thomas More chose to situate Utopia on an island, extremely well-guarded from the outside world (More, 2007: 33), whilst John Rawls strives to formulate 'a reasonable conception of justice for the basic structure of society conceived for the time being as a closed system isolated from other societies' (Rawls, 1999: 7).[2]

Both the work of John Rawls and many Utopian scholars share the belief that the social transition toward some ideal requires a conception of that ideal. But, the work of John Rawls differs from the Utopian theory of Mannheim, Bloch and Goodwin in that it does not explicitly discuss the relationship of its *object* (in Rawls: social justice) to Utopia. However, if the concept of Utopia can be integrated into Ideal Theory it becomes much more comparable to the works mentioned above as utopian. With this, the compatibility of the different works on ideal societies can be examined, opening up the possibility of a wider, cross-discipline (or cross-object) account of the theory of ideal societies. Thus, as we have seen these scholars profess the importance of the concept Utopia to their respective theories, we can suggest that Utopia could be of similar importance for (ideal) theories of justice. At the very least it may serve to highlight the parallels that are currently overlooked by ignoring Utopia.

Now we can summarize the initial benefits of viewing Ideal Theory as a Utopian Theory. Firstly, the inclusion of the concept of Utopia serves to remind us that Ideal Theory should not be limited to theories of justice alone. Utopia's most often advocate some specific virtue(s), hence; there can also be Ideal Theories of Morality, Happiness, Freedom, etc.[3] Consequently, we should refer to Rawlsian Ideal Theory as an Ideal Theory of Justice. This frees the term Ideal Theory up to be used with reference to other disciplines. Secondly, the incorporation of Utopia emphasises

that the Rawlsian definition of Ideal Theory is a narrow one; Ideal Theory should do more than achieve a reasonably good society, relatively free from injustice. The preceding two reasons demonstrate the most important benefit of introducing Utopia to Ideal Theory: it serves as conceptual clarification. Furthermore, referring to a Utopia as the product of Ideal Theory, and thus of Ideal Theories of Justice, underlines the fact that in Rawlsian Ideal Theory: (1) some favourable conditions are assumed (this is arguably true of all Utopia's), (2) full-compliance is a condition and (3) that there can be no instances of societal injustice. This then makes Ideal Theories of Justice utopian. Additionally, Rawlsian Ideal Theory can benefit from the warning of history often associated with Utopia in our post-communist world.

Thus far I have re-examined Rawlsian Ideal Theory, sketched its parallels with Utopian Theory and concluded that the use of Utopia benefits its conceptual clarity. Now we proceed to the second aim of this paper, which is to examine the consequences of this new point of view on two articles from contemporary literature regarding Rawlsian Ideal Theory. The first article defends Ideal Theory, whilst the second stands in opposition.

Part II

Zofia Stemplowska: *What's Ideal About Ideal Theory?*

Zofia Stemplowska opens her paper by remarking that Ideal Theories 'share a common characteristic: much of what they say offers no immediate or workable solution to any of the problems our societies face' (Stemplowska, 2008: 319). She does not, however, consider this a fatal flaw and sets out to defend these theories by contending that the debate regarding Ideal and non-Ideal theories can be productive, if they are not treated as rival approaches to political theory. To this end she offers her own definition of Ideal Theory by examining the structure of normative theory. Initially she writes:

> 'One crucial difference between various normative theories concerns whether they offer viable recommendations, where by viable recommendations I mean recommendations that are both achievable and desirable' (Stemplowska, 2008: 324).

Stemplowska calls these 'AD-recommendations' and believes that it is the absence of these recommendations that is crucial in separating Ideal from non-Ideal Theories. Consequently, Stemplowska defines non-ideal theory as 'theory that issues AD-recommendations, and ideal theory as theory that does not' (Stemplowska, 2008: 324). With this definition she goes on to say that normative theories may lack these AD-recommendations for different reasons. Firstly, they may offer recommendations that cannot be considered AD-recommendations, and secondly they may not aim at offering any such recommendations at all. She claims that the latter serve only to clarify our understanding of certain values and principles, and can therefore not be objectionable. Of the former, she proceeds to identify several further sub-categories. Ignoring what she calls 'bad theories', Stemplowska identifies '(a) theories that fail to issue AD-recommendations because they ignore the fact of non-marginal noncompliance, and (b) theories that fail to issue AD-recommendations because, even with full compliance, there is no solution to the problem for which recommendations are sought' (Stemplowska, 2008: 331).

In what follows, Stemplowska defends theories that do not offer AD-recommendations, concluding that they are nevertheless indispensable to normative theory. Furthermore, she claims that accepting her definition of the distinction between Ideal and non-Ideal Theory allows us to see that complex normative theories are likely to contain both Ideal and non-Ideal aspects.

Although Stemplowska defends Ideal Theory, her contention that the identification of AD-recommendations best resolves what is at stake in the debate between Ideal and non-Ideal Theories fails to be convincing. The main problem with her approach is that she offers a negative definition of Ideal Theory, meaning she defines Ideal Theory by what it is not. She identifies a characteristic that is most often associated with non-Ideal Theory and then attempts to define and categorize Ideal Theories by virtue

of the absence of this characteristic. This results in the need to categorise different Ideal Theories according to why they do not meet the criterion of supplying AD-recommendations. With each of these additional categories come further issues of definition. The obvious problem with this approach is that it is unclear how these divisions are to be made.

Furthermore, Stemplowska approaches the definition of the distinction between Ideal and non-Ideal Theory from the bottom-up. Meaning, she takes a characteristic of the majority and defines the minority by its absence. Consequently, her definition of non-Ideal Theory is too broad. Undoubtedly there is a much greater body of non-Ideal Theory, but using a common feature of the majority as the basis for a negative definition of the minority may lead to several undesirable consequences: (1) Ideal Theories may falsely be labelled non-Ideal; some of Rawls' recommendations are arguably both achievable and desirable, for example, and (2) non-Ideal theories may be labelled ideal, as their output could be argued to be both unachievable and/or undesirable. It would be more fruitful to characterise Ideal Theories by virtue of some unique feature, as opposed to the absence of a common feature. In other words, what is required is a top-down approach. The amended definition of Ideal Theory with reference to Utopia is an example of this top-down method for identifying Ideal Theories. Defining Ideal Theory as a system of principles that, when fully-complied to by all members of society, results in a Utopia of social justice, would preclude the need for the problematic sub-categories of Ideal Theory described by Stemplowska. The image of the top-down approach thus captures the Rawlsian idea that some Ideal Theory is required primary to non-Ideal theorising. Moreover, the addition of Utopia to the definition of Ideal Theory also serves as a *necessary* condition of Ideal Theory, as any non-Ideal Theory cannot result in Utopia.

Although Stemplowska's proposal faces problems that are likely insurmountable, she succeeds in highlighting a poignant difference between Ideal and non-Ideal theory in general. Unfortunately, the resulting attempt to distinguish between Ideal and non-Ideal Theory based on this difference is highly problematic and therefore not useful as a tool for differentiation. Besides lacking the theoretical virtue of being narrow, the problems affecting Stemplowska's proposal suggest that an approach singling out a common positive feature of Ideal Theory is preferable.

Charles W. Mills: *"Ideal Theory" as Ideology*

Having examined an article that accepts the need for Ideal Theory, it is important to discuss another that rejects it. This will show us if the new understanding of Ideal Theory can survive established criticisms. Charles W. Mills proves himself to be a vocal opponent of Ideal Theory in his *"Ideal Theory" as Ideology*. His article is nothing short of an all-out attack on Ideal Theorizing. He sets out not only to discredit Ideal Theory, but to prove that non-Ideal Theory is superior in every way; going so far as to say that even the act of engaging in Ideal theorizing perpetuates the non-ideal (Mills, 2005: 182). Mills proposes that only Non-Ideal theorizing can offer solutions to the non-ideal. To serve his ends, Mills employs issues such as gender and race inequalities to demonstrate the need for non-Ideal Theory. Throughout his article, Mills offers possible definitions of Ideal Theory and argues why these do not hold. I wish to show that on two occasions Mills mistakenly dismisses Ideal Theories.

Mills begins by distinguishing different types of theorizing of which the most important, in this context, are ideal-as-idealized and ideal-as-descriptive. The former referring to an idealized model of what some ideal *P* should be like, the latter being a somewhat idealized or abstracted model of how *P* actually works. Mills then builds on these to define Ideal Theory, he writes: 'What distinguishes ideal theory is the reliance on idealization to the exclusion, or at least the marginalization, of the actual' (Mills, 2005: 168). This is of course not a strict definition of Ideal Theory for it is difficult to determine the extent of reliance on idealization, let alone the marginalization of the actual or even what the actual may be. However, Mills does go on to specify Ideal Theory further: 'ideal theory either tacitly represents the actual as a simple deviation from the ideal, not worth theorizing in its own right, or claims that starting from the ideal is at least the best way of realising it' (Mills, 2005: 168). Firstly, I do not believe that any substantive Ideal Theory tacitly represents the actual as a *simple* deviation from the ideal. Secondly, none hold that this is not worth theorising in its own right. According to Mills then, Ideal Theories must then claim that

starting from the ideal is the best way of realizing the ideal, a claim which he believes to be false.

Mills defends this claim by quoting John Rawls: 'The reason for beginning with ideal theory is that it provides, I believe, the only basis for the systematic grasp of these more pressing problems' (Rawls, 1999: 8). Mills mistakenly equates his own claim, that in the view of Ideal theorists 'starting from the ideal is the best way of realizing it', with Rawls' statement. He compounds this mistake by later adding: 'the argument has to be, as in the quote from Rawls above, that this is the *best* way of doing normative theory, better than all the other contenders' (Mills, 2005: 171 – italics in original). To start, Rawls certainly does not claim that Ideal Theorizing is the best way to realize the ideal, only that it is a necessary component of this process. Additionally, he makes no claim to have discovered 'the *best* way of doing normative theory'. What Mills points out at best is that most, if not all, Ideal theorists claim that Ideal Theory is a necessary step toward realizing the ideal; not the best nor the only.

Amongst Mills' many criticisms there are two in particular that illuminate the error in his dismissal of Ideal Theory. In a lengthy section on *The Vices of Ideal Theory*, Mills wishes to 'quickly clear away some of the ambiguities and verbal confusions that might mistakenly lead one to support ideal theory' (Mills, 2005: 170). Of these 'verbal confusions' the first conception of Ideal Theory that is mistakenly dismissed by Mills is that of Ideal Theory being 'just a model' (Mills, 2005: 171). It can be defended that Ideal Theories are just that: theories. As has been said, Rawls was satisfied with the possibility of formulating a reasonable conception of justice for the basic structure of society (Rawls, 1999: 7). On the other hand, far from this being *just* a theory, any reasonably successful attempt to model an ideal society based on a set of principles would be no small feat, and would be invaluable to the study of political and social theory in philosophy. Moreover, an Ideal Theory need not say anything about the non-Ideal, nor offer any value judgements or achievable and desirable recommendations for that matter. It could function as 'just a model', placing the burden of implementation on non-Ideal Theories. Ideal Theory may serve only to demonstrate or test the compatibility of certain ideals proposed in a theory. If these professed ideals prove to be compatible, the Ideal Theory would be the blueprint for Utopia.

The second misconception reads 'Nor does the simple appeal *to* an ideal (say, the picture of an ideally just society) necessarily make the theory ideal theory, since nonideal theory can and does appeal to an ideal also' (Mills, 2005: 171 – italics in original). However, I would argue the exact opposite. Appealing to a 'picture of an ideal society', or Utopia, *does* necessarily make the theory Ideal Theory. The fact that a paradoxical assumption of favourable conditions is made, the fact that there must be full-compliance to the prescribed ideals, the fact that there can be no instances of injustice and the fact that these conditions constitute the ideal state, or Utopia, make such theories Ideal Theories. The inclusion of Utopia only serves to *underline* that these conditions are part of Ideal Theory. Moreover, the appeal of any non-Ideal Theory to some ideal would require some conception of that ideal, which in turn requires some theory of said ideal. It is at this point that non-Ideal Theory necessarily appeals to Ideal Theory. In other words, Mills was wrong to reject Ideal Theory on the basis of these arguments. Moreover, we can better understand why Mills fails to reject Ideal Theory if we refer to Ideal Theory with reference to Utopia.

In conclusion, the addition of Utopia to the conception of Ideal Theory helps save it from the Mills' criticism by highlighting where he falsely dismisses Ideal Theory. Mills does not succeed in relegating Ideal Theory to a sub-par status in social and political philosophy. In fact, if Mills were to advocate some version of Ideal Theory it would be one incorporating Utopia, for with Utopia come the many warnings of history against blindly implementing ideology. He would fervently endorse the historic dimension that Utopia brings to Ideal Theory as he himself so often appeals to it.

Conclusion

In this paper I initially re-examined the Rawlsian definition of Ideal Theory and explored the relationship between Ideal Theory and Utopian Theory. After finding structural similarities, I introduced the concept of Utopia to Ideal Theory of Justice, resulting in a new definition of Ideal Theory. A complete Ideal Theory of Justice would be: a system of principles that, when fully-complied to by all members of society, results in a Utopia of

social justice. I found support for this new definition by examining Zofia Stemplowska's article on Ideal Theory, which demonstrated the need for a positive definition of Ideal Theory, meaning that it must identify some characteristic of Ideal Theory that non-Ideal theories lack, such as a conception of Utopia, as opposed to being labelled Ideal by virtue of lacking some feature that non-Ideal theories share. Additionally, the new conception of Ideal Theory withstands attempts by Charles W. Mills to render all Ideal Theory irrelevant.

To conclude, I hope to have shown that a definition of Ideal Theory with reference to the ideal society Utopia, is not only possible, but desirable; for it can both clarify the debate and withstand significant criticism. With this, we may be encouraged to look more closely at the similarities between Utopianism and Ideal Theory.

Volker Ruitinga (1987) completed his Bachelor's degree in Philosophy at the Erasmus University, and is currently enrolled in the Master of Philosophy. Over the course of his studies he developed an interest in Utopia, and is currently writing his master thesis on the concept of Utopia and Ernst Bloch's 'Das Prinzip Hoffnung'.

'Ideal Theories and Utopia' was written for the completion of the mastercourse 'Contemporary Theories of Justice' by prof. dr. Ingrid Robeyns.

Notes

1 Elaboration on the reasons for this conviction are beyond the scope of this essay but rely generally on the idea that to change the world for the better requires some conception of what this better world would be.

2 Moreover, it is no coincidence that both More and Rawls introduce their works by citing these conditions.

3 An Ideal Theory of Justice envisions a Utopia free from injustice, whereas an Ideal Theory of Freedom, for example, would conceptualize a society whose members suffer the bare minimum of constraints on their actions.

Literature

Bloch, E. (1985) *Das Prinzip Hoffnung*. Frankfurt am Main: Suhrkamp.

Claeys, G. (2010) 'Preface'. In G. Claeys (ed.) *The Cambridge Companion to Utopian Literature*. Cambridge: Cambridge University Press.

Davis, J (1981) *Utopia and the Ideal Society*. Cambridge: Cambridge University Press.

Goodwin, B. & Taylor, K. (2009) *The Politics of Utopia*. Bern: Peter Lang.

Levitas, R. (2011) *The Concept of Utopia*. Oxford: Peter Lang.

Manheim, K. (1954) *Ideology and Utopia*. New York: Harcourt, Brace.

Mills, C.W. (2005) '"Ideal Theory" as Ideology'. In: *Hypatia* 20, No.3.

More, T. (2007) *Utopia*. Sioux Falls: NuVision Publications.

Rawls, J. (1999) *A Theory of Justice: Revised Edition*. Harvard: Harvard University Press.

Robeyns, I. (2008) 'Ideal Theory in Theory and Practice'. In: *Social Theory and Practice*, Vol.34, No.3.

Stemplowska, Z. (2008) 'What is Ideal About Ideal Theory?'. In: *Social Theory and Practice*, Vol.34, No. 3.

Vieira, F. (2010) 'The concept of utopia'. In: G. Clayeys (ed.) *The Cambridge Companion to Utopian Literature*. Cambridge University Press.