

HIER BEN IK

HIER BEN IK

Liber amicorum voor Monica Meijsing

Hans Dooremalen

redactie

Tilburg, 2020

Departement Filosofie Tilburg University

© diverse auteurs, augustus 2020

Het auteursrecht ligt per essay bij de auteur of auteurs ervan.

Omslagontwerp & vormgeving: Hans Dooremalen.

Brave New Books

Departement Filosofie, Tilburg University

ISBN: 9789464050875

Persoon-zijn is een status, een soort eretitel,
die we *aan elkaar en onszelf* toekennen.

— Monica Meijsing

INHOUD

VOORWOORD

Hans Dooremalen — 11

ALFRED ARCHER, MAUREEN SIE & NATHAN WILDMAN

Personhood in the digital age

An interview with Monica Meijsing — 13

TON DERKSEN

Twee keer levenslang, twee keer wetenschapsfilosofische fouten — 31

JEANNE PEIJNENBURG

Een goed leven?

Retro-causaliteit en het tweedepersoonsperspectief — 51

WILLEM B. DREES

Daar ben je!

Godsdienstfilosofische kanttekeningen — 69

KATRIEN SCHAUBROECK

Zelfbewuste zebra's en kinderen die naar koekjes verlangen

Komt het zelfbewustzijn in gradaties? — 85

MACHTELD GEUSKENS & MATTEO COLOMBO

Breaching taboos at work

A dialogue on (while) breastfeeding — 103

JOACHIM LEILICH

Taalspelen en categoriefouten
Over Wittgenstein en Ryle — 123

HERMAN DE REGT

Ooit ben ik nergens
Een reactie op Monica Meijssings idee van een persoon — 137

INGEBORG VAN DEN BOLD

Papieren Otto's en personen — 171

RUUD WELTEN

Waar ben ik als ik er niet ben? — 187

LEON DE BRUIN

Ben ik een brein in een vat? — 203

FILIP BUEKENS

Over patrijzen en mussen
David Hume en de waarde van waarheid en kennis — 215

WILLEM VAN DER DEIJL & HUUB BROUWER

Wat voor iemand moet je zijn om mee te tellen? — 229

SANNEKE DE HAAN

Jezelf zijn doe je niet alleen — 245

JENNY SLATMAN

Komt een ‘persoon’ bij de dokter — 257

MARC SLORS

Leve het lichaam

Meijsing en Schechtman over het leven van personen — 275

SANDER VERHAEGH

Personen, walvissen en genderneutrale paspoorten

Een analyse van Meijsings classificatiefilosofie — 293

WIM DUBBINK

Ik ben door te moeten, of ik ben niet

Een kantiaans antwoord op *Waar was ik toen ik er niet was?* — 305

HANS DOOREMALEN

Is het wat om een kat te zijn? — 359

Over de auteurs — 385

VOORWOORD

In de vroege ochtend van 24 mei 1987 reed Kenneth Parks twintig kilometer van zijn eigen huis in Pickering naar dat van zijn schoonouders in Scarborough. Hij vermoordde zijn schoonmoeder met een wielmoersleutel en probeerde zijn schoonvader te wurgen. Hij deed dit – zo beweerde hij later – allemaal in zijn slaap. Hoewel men hem in eerste instantie natuurlijk niet geloofde, werd dit na wetenschappelijk onderzoek toch waarschijnlijk geacht. Parks werd vrijgesproken: Hij was er blijkbaar niet bij toen hij dit ongewenste gedrag vertoonde. Waar was Parks toen hij zijn schoonmoeder vermoordde? Kan je echt dingen doen zonder dat je er zelf bij bent? Om die vragen te kunnen beantwoorden, moet je weten wat een persoon is.

In haar boek *Waar Was Ik Toen Ik Er Niet Was?* pakt Monica Meijnsing de vraag naar wat een persoon is op. Met de hulp van denkers uit het verleden en het heden formuleert ze een origineel antwoord dat eenieder die zich met deze kwestie bezighoudt zou moeten lezen, overdenken en doorvertellen.

Het beantwoorden van de vraag naar wat een persoon is, vereist reflectie op andere kwesties: Een persoon is een bewust wezen, dus moeten we weten wat bewustzijn is. Een persoon heeft een brein, dus moeten we nadenken over de vraag of het voldoende is om een biologisch brein te hebben om een persoon te kunnen zijn en over de vraag wat de rol van het lichaam is. Wat kunnen we leren van de wetenschap als het hierom gaat En wat kunnen we leren van filosofen als Descartes, Locke, Hume, Ryle, Rorty en Kant? Waarom

willen we eigenlijk weten wat een persoon is? Personen houden we verantwoordelijk voor hun daden. En dat houdt in dat we voorzichtig moeten zijn met hoe we met personen omgaan: In dit boek wordt door enkele van Monica's academische vrienden gereflecteerd op een keur aan kwesties inzake de vraag wat is een persoon?

Ik ben veel dank verschuldigd aan Marie Stolk voor de redactionele hulp bij de totstandkoming van dit boek. Marijke Nitsche dank ik voor de organisatorische ondersteuning bij dit project. Ook wil ik graag Florian Jacobs en Marthe Kerkwijk van de Internationale School voor Wijsbegeerte danken, die in het complot zaten om Monica zonder dat ze het wist een bijdrage voor haar eigen *liber amicorum* te laten leveren.

Met elke bijdrage in deze bundel leveren de auteurs inzichten die tot verdere vragen leiden. Met elke bijdrage in deze bundel brengen de auteurs hulde aan Monica bij haar afscheid van het Departement Wijsbegeerte van Tilburg University. Met elke bijdrage zegt de auteur ook tegen Monica: 'Hier ben ik.'

— Hans Dooremalen
Tilburg, augustus 2020

PERSONHOOD IN THE DIGITAL AGE
AN INTERVIEW WITH MONICA MEIJSING

Alfred Archer, Maureen Sie & Nathan Wildman

Monica Meijsing's book *Waar Was Ik Toen Ik Er Niet Was* (*Where Was I When I Wasn't There*) was published in 2018. It opens with an anecdote very recognizable for those who know Meijsing's dry and quick sense of humor. After waking up from surgery a nurse tells her it has been quite a bloody affair, making a new nurse faint on the spot, to which Meijsing instantly replies with a 'glad I wasn't there then.' She continues by unpacking the joke, explaining why the 'I wasn't there' claim at once sounds quite true and strange.¹ In doing so the reader is immediately plunged deep into a philosophical discussion. How should we understand the claim that there was a 'black hole' *in* my consciousness, when I am under narcosis? Is that a claim we can make sense of? After all, I *wasn't* conscious during the anesthetic period; but I *was* conscious of going into narcosis and waking up right after it as well. In order to make sense of talk about a black hole *in* my consciousness, it needs to make sense to claim that there is an 'I' when I am under general anesthetic, when there is no consciousness, when there is no first person perspective.² The claim that I wasn't there when under narcosis sounds true made from the first person

¹ Meijsing (2018): 11.

² Meijsing (2018): 12.

perspective, but strange from the so called third person perspective, the perspective that looks at someone from the outside, noting that, clearly, the person does not disappear even when she is under narcosis.³

To answer the question where I was, during narcosis, Meijnsing points out, we need to answer the question *what* am I: a body, a perspective on the world, a consciousness? After we have answered that question we need to reflect on the conditions under which I persist, and how to think of *where* we are and *whether* we are, during narcosis. Her answer takes us through a number of philosophical discussions, about how consciousness, personhood and the concept of ‘a self’ relate to each other, how all of those topics relate to our existence as bodily beings exactly, how to think about the persistence conditions of persons, how our nature as persons relates to our nature as embodied and acting agents and how human consciousness comes into being. We warmly recommend the book to everyone who speaks Dutch and hope it will be translated into English soon, for those who do not. Not only does it take you through the philosophical discussions on this topic and illustrate these discussions vividly with many interesting case studies from neurology, neuroscience and psychology, it also articulates Meijnsing’s own view. It is this view – the basics of which she will introduce herself – which made us wonder what her thoughts were on some issues the internet, virtual reality, social media and AI raise for our self-perception and our ideas on personhood. Rather than figure this out for ourselves, we decided to ask her. We hope you enjoy.

INTRODUCTION

First of all, great that you are prepared to do this – we are really looking forward to this interview with you on the topic of personhood in the digital age. Could you, by way of opening, tell us a bit more about your view on persons and personhood?

³ Meijnsing (2018): 13.

You are entirely welcome – the pleasure is mine! I really appreciate discussing these themes with you.

My position is that you have to distinguish two quite different questions: the first question is what kind of beings *we*, human beings, are. And the second is what a person is. The first question is metaphysical, and asks for a metaphysical answer. What I am trying to do is a kind of descriptive metaphysics. Put in a nutshell: I argue against both body-mind dualism and physicalism⁴ and for a dual aspect or neutral monist⁵ kind of metaphysics of the human being. But – and this is something a bit more controversial – the second question is not metaphysical at all. I argue against the idea that there is some metaphysically real category of persons distinct from the category of human beings or human organisms. Instead I opt for a relational, and completely human-dependent notion of a person. The practice of accepting certain beings into our already existing community of persons is prior to any characterisation of what a person is. That implies that any characterisation of what a person is follows from this practice, and is shifting across time and cultures, just as these practices are shifting.

I myself find this a scary position, because there is nothing that grounds these practices, and that seems to imply a thoroughgoing relativism of what personhood is. Yet I also want to believe that we can do this accepting in ways that are normatively evaluable: that we can be *wrong* in certain practices and should strive to get it *right*. That seems to commit me to a kind of moral realism, and I know that that is a position fraught with difficulties. But, as Maureen once told me, I am not obliged to solve all philosophical problems (pew!).

⁴ The idea that there only exist physical things, i.e., that if a mind exists, it is not made of a different substance.

⁵ Monism is the term for the theory that everything is made of the *same* substance, most often believed to be physical matter. A dual aspect or a neutral monist argues that there is only one substance and that there are two ways of viewing or describing this one substance, i.e., in mental or in physical terms.

So yes, I do believe that nothing *grounds* the notion of personhood, and that it is up to us, the community of persons, to do the right thing by contenders for personhood.

ROBOT PERSONS

If we understand it correctly your view on personhood—on what makes us persons—differs in several respects from other philosophical views. One of the respects in which it differs is that according to you there is not one unique characteristic that distinguishes persons from non-persons, such as, for example ‘a soul’ or ‘rationality.’ As you outline in your book, the distinction between persons and non-persons comes into existence gradually and in practice: by being treated as persons by other persons.⁶ Now you also claim that what it is to be ‘treated as persons’ is something complex and very practical as well: it entails being treated as beings with a first person perspective, as autonomous, and as moral beings, that is, beings that are a proper subject of the moral reactive attitudes such as resentment and blame, gratitude and praise.⁷ Hence, we ‘inherit’ our personhood of those who treated us like persons and who took care of us.

In your book it is clear that you articulate this view to deal with certain philosophical issues. But it seems that it enables us to be more liberal than most are generally inclined to be with respect to the question of what entities in our world could become persons or could become persons to a certain degree at the very least. This would include, for example, robots, especially those robots that look and function like us. Do you agree with this, and if so, what would you say to persons that worry about the clear absence of a soul or freedom of will that such machines would have?

You don't ask simple questions, do you 😊?

⁶ Meijnsing (2018): 234.

⁷ Meijnsing (2018): 236. Reactive attitudes are attitudes with which we respond to other people's behavior. They were made influential in the discussion about moral responsibility by Peter F. Strawson.

First two disclaimers:

- I don't do predictions.
- If robots are indistinguishable from us, they are indistinguishable from us.

Elaborations:

If it is true that personhood is non-metaphysical and relational, and dependent on our practices, it means that inclusion in the community (or commonwealth) of persons will not remain as it is today, just as it is different today from what it was in the past. So my non-metaphysical notion of personhood *implies* that I cannot say what will happen in the future – quite apart from the fact that predictions about culture and human beings are always tricky and seldom come true.

The other part I find more interesting, or rather, I can say more about it. My PhD thesis was titled *Mens of Machine? (Human Being or Machine?)*, and I used to argue that programmed computers were quite different from human beings (you may be surprised how controversial that was back then in the AI community!).⁸ People used to say to me: 'But if we could build a machine that was completely indistinguishable from us, how could you say that it still was different from human beings?' And of course, I could not say that under that premise. The really interesting question is: 'What do you mean by *indistinguishable*?' If you mean 'the same in absolutely every respect', then of course, by definition, they are not different in any respect. What you have to answer is: 'How many respects, and which respects, have to be the same in order to accept a machine in the community of persons?'

⁸ The, currently disreputed, John Searle had made the same claim in his article 'Minds, Brains, and Programs' (1980), but almost all reactions to his article were very negative: according to his critics, among whom Dennett was one of the most vociferous, Searle simply hadn't understood computers!

Your question is about ‘robots that will look and function like us.’ Let’s first concentrate on the looks – that is of course the first thing we notice about another being. But how important is it for inclusion as person? In general, if you want to be inclusive, and if you want to be against racism and lookism, looks should not matter (sadly, they always do). But seriously, where should we draw the line between ‘one of us’ and ‘alien’? I think that interaction, affective resonance, having and noticing reactive attitudes, all of that is more important than looks.

Let me give you some real life examples: over the past years I have been present at a number of death beds. Most of the dying persons were no longer conscious, and interaction was no longer really possible. Yet it felt like something, and it raised reactive attitudes in me, to hold their hands. What struck me forcibly, was how *entirely different* they looked almost immediately after dying. Some of these people were blood relatives of me, so they looked, objectively, a lot like me. Yet, after they died, they didn’t *feel* like me at all. I have more affective resonance with my cats or my horse (and I don’t look much like either of them) than with these dead people. So how important is ‘looking like us’?

Yet maybe robots have to look like us in some very specific respects in order to spark reactive attitudes in us, and in order for us to notice reactive attitudes in them.

Maybe what’s important is that they must seem to be *alive*. ET doesn’t look much like a human being, but the fact that he breathes, seems to have a heart, and has a face with eyes and different facial expressions, facial expressions that seem to disclose an emotional life, makes us *feel* for and with him.

Which brings me to ‘functioning like us’. How much like us a robot has to be in order to function like us? And now I am going rather quickly and

roughshod over a difficult terrain: can a robot function like us if it isn't alive? Which means, if it isn't an autopoietic system?⁹ Aren't reactive attitudes intimately connected with being a precarious system that constantly has to strive for self-preservation? Could we see reactive attitudes in and feel them for a being that is not in this way autopoietic, and moreover, warm and fragile (we don't seem to connect very much with cold-blooded creatures; it is actually also an expression for human beings that are somewhat less of a person!). Could a robot, which you can turn off and turn on again at an arbitrary later time, ever be autonomous?

Concluding: yes, if a robot functions like us in all or most of the above-mentioned respects, and therefore looks like us at least in some respects, we might one day include it in the community of persons (but who am I to predict).

What you sketch seems most certainly to be what the fiction on robots is struggling with! If you watch the popular HBO series Westworld, for example, it seems to be this striving for self-preservation that is connected to the fragility of the robot-inhabitants; and we also see it very clearly in Asimov's wonderful Bicentennial Man. In the recently published Machines Like Me by Ian McEwan, one of the major events in the book is that the main character, a robot called Adam, disables the button with which he can be turned off. So, it seems that those features you highlight – of fragility, warmth, striving for self-preservation – seem intimately connected to autonomy and to the susceptibility for reactive attitudes. What about the soul and free will? How do these figure in this web?

I don't know what souls are; I don't know if I have one. I am a monist, so I don't believe in souls in a dualist interpretation. I do use the word 'soul' in

⁹ As Meijnsing explains in her book: 'Autopoiesis is the Greek word voor self-making, self-construction. An autopoietic system is [... one] that keeps together its own organism. That is to say, the system is built in a way that it keeps its internal system constant and does so within strict boundaries, in changing environments.' See: Meijnsing (2018): 204.

everyday life to refer to what is important to us, to our emotional and affective states, to our innermost and crucial values. But as long as a robot has the same kind of things that matter to it, emotional and affective states, and innermost and crucial values, why should a soul be clearly absent?

I forgot a third disclaimer: I don't do free will. I mean, I find this problem of free will totally insolvable. I don't know what it takes to have or not have free will, and I don't know whether free will is possible at all (but I think it is). So I won't go into that discussion (though I think that Peter Strawson, in his 'Freedom and Resentment', has said very wise and true things about the topic).

HUMAN BODIES AND MEDIATION

In your book you also talk about the importance of our embodiment, the fact that we are the kind of organisms that we are and that we have the bodies that we have. For example, you insist that it is the human organism that has what you call a first person point of view, and that we build from the more primitive first perspectives on our surroundings, from an early age onwards. In our contemporary society, the digital age, we see that children interact with technology from an increasingly earlier age. Also this interaction, for many, persists into later in life and it is not uncommon to see people immersed in interaction with their virtual worlds at the expense of their interaction with the non-virtual world.

Does it follow from your theory/ do you think that this impacts their development in ways that will make us into a different kind of persons [in any fundamental sense]?

Yes, I do. I am very worried that young people nowadays seem to have less bodily interaction with others, and less eye contact, than they used to do. Just as the excessive use of screens impacts their eyesight – the incidence of nearsightedness has grown explosively – excessive interaction with non-

embodied beings may impact their being a person. If reactive attitudes and the first-person point of view are crucially dependent on being embodied and being surrounded by embodied others, then in the absence of other embodied beings, becoming a person may be that much more difficult. Or, alternatively, it might be that the whole concept of a person will change quite radically, and in ways that I find worrying.

But then, as long as older people like myself have said anything about ‘today’s youth’ at all, they have said that it was going to the dogs, so what else is new? And again, who am I to predict?

AVATAR IDENTIFICATION

Psychologists are increasingly interested in the way in which our interaction with virtual worlds influences our behavior. One interesting finding, called the Proteus Effect, is that the way in which people behave in online environments is shaped by the appearance of their avatar. In one study, for example, players who were given a taller avatar were found to behave more confidently in negotiations than players assigned a shorter avatar. In addition, interacting with different avatars in virtual worlds has also been found to shape how people behave outside of that world. Do you think these findings have anything interesting to tell us about the nature of personhood?

I didn’t know about the Proteus Effect. It sounds quite plausible, but I have become a little wary of social psychology findings – so many of them, among which are some of the most influential and famous, turn out not to be replicable. But that’s maybe not relevant.

What I think is the case, is that people always have fashioned themselves to a certain extent: *The Presentation of Self in Everyday Life* by Ervin Goffman (1956) was one of the first books to pay extensive attention to this phenomenon, but of course Machiavelli’s *Il Principe* is a much earlier manual

of how to fashion a public self. I also think it is quite obvious that we present different versions of ourselves (or our selves) in different social circumstances. So the fact that the use of avatars interact with these self-fashionings and self-presentations, consciously or unconsciously, doesn't seem to me very surprising, or indeed a new phenomenon.

So the phenomenon of self-presentation has a lot to do with what we might call *personality*, or even with the self. It has a lot to do with having different, and sometimes even rather incompatible, sides to ourselves (or our selves; I keep putting this in parentheses, because I think the very notion of *a* self, as a separate entity, is extremely problematical, but of course we keep talking about it anyway).

It is a different matter how much this phenomenon has to do with the notion of personhood, and I am not sure whether this phenomenon has much to tell us about that. People who interact with virtual worlds and play with avatars usually have been persons for years. I think the process of becoming of a person is something which starts from birth and is, in a way, finished, long before adulthood. Different social groups have been mirroring a slightly different kind of person for the individual since birth as well. This accounts for the many-sidedness and multi-dimensionality of most personalities. I don't think it changes very much for the process of becoming a person as such.

SOCIAL MEDIA AND HUMAN INTERACTION

Another way in which the digital age may influence human behavior is through the sheer number of potential human interactions that are opened up to us by social media. Someone posting something controversial on Twitter, for example, may quickly be subjected to hundreds of responses from strangers all over the world. Similarly, previously private moments such as eating one's dinner can now be instantly shared online. This leads to two questions. First, do you think the number of people we can instantly interact with may influence the way we

interact with one other? Second, do you think that opening up more and more of our lives to human interaction may influence how we develop personhood?

Being almost digitally illiterate (I think I was among the first people who used e-mail, back in the early eighties, but I do not use social media except for the occasional WhatsApp message, where I have some twenty contacts) I am not very confident with answering this question. I simply cannot understand why people should actually choose to share so much of their life with countless others – I think my private life is my own, and I try to remain as invisible as possible in this world.

But are people really interacting with so many contacts? As a firm believer in the importance of embodiment – in fact, I would rather say that we are animated or ‘ensouled’ bodies than that we are embodied souls or minds – I think that nothing can substitute for real, bodily, face-to-face interactions. I find it worrying to read that young people do not use phone conversations, but prefer text conversations because phone conversations are too direct. If that is true, it seems to suggest that, because of social media, people are not interacting more, they are interacting less. Phone conversations are hugely impoverished compared with real, face-to-face conversations, but at least some paralinguistic characteristics of what someone is saying get conveyed. In text conversations even these are lost – and emoticons are definitely not an adequate substitute. Shared affectivity is that much harder to achieve.

I believe that the being included in the community of persons starts with shared intentionality and shared affectivity, and for that you need a body that reacts quite directly to other similar bodies, as indeed new-born infants already react to and interact with human beings in a way they do not react to and interact with non-human beings and entities.

But again: by the time people start using social media, they have already begun being part of the community of persons. So maybe their use does not influence how we develop personhood. Of course there *is* influence in another way: the tendency of parents and care-takers and generally *other* persons to be more absorbed in their screens than in their actual surroundings, which includes persons-to-be, and the tendency of some parents to mostly take pictures of their children and look at them, instead of at the children themselves, *does* influence how these persons-to-be develop, I guess.

Thanks so much for those answers, Monica. For us, as well as provoking more interesting questions about the digital age we are living in, they also invite some questions that bring us back to your first sketch of your view on what it means to be 'us' (what kind of beings we humans are) and what it means to be a person. Maybe it is nice to close with those questions, although please feel free to refer us to parts of your book!

In your answer above you say that: 'Being included in the community of persons starts with shared intentionality and shared affectivity...', which sounds interesting and plausible. However, given the fundamentally relational nature of personhood on your view, someone might worry that there is a circularity concerning what it means to be a person. You seem to claim that we become persons because other people (persons) treat and interact with us as persons, that they depend upon – an 'already existing community of persons', as you said earlier. But does that not mean that there is no way for persons to ever come into being? Suppose that Eve is 'the first person.' For Eve to become a person, there must already exist some community of persons that she is embedded in. However, if this community of persons already exists, then Eve is not the initial person. Of course, Eve is just arbitrary – the same argument will apply to any individual that we pick as the 'first' person. So, do you think there is a way to get around this problem, e.g. by having persons 'bootstrap' their way into being?

Lastly, we wanted to press you some more on the relationship between relativism about personhood and your idea that we should ‘do the right thing by contenders for personhood’. Suppose far into the future and robots have come into existence and have acquired personhood and that, for whatever reason, the person community split into two isolated groups, which could not communicate with each other (see for a fictional example Westworld, the series produced by HBO – but if you do not like hypothetical SF scenarios you can just use another example of communities getting isolated from one another).

Furthermore, over time, these communities evolve in a variety of ways, changing and modifying who they ascribe personhood to.

Finally, imagine they encounter each other again. Due to their divergent cultural/ social evolution, neither is willing to label members of the other group persons. Given your view, would this sort of scenario entail that there were two distinct types of person – i.e., person₁, defined in terms of one community, and person₂, defined in terms of the other? And, more importantly, in refusing to confer personhood to these others, are these communities making a mistake? If so, is it a metaphysical one or an ethical one (or something else altogether)?

Thanks again for these intriguing questions. I’ll deal with them in the same order.

So what about Eve? This is of course the type of chicken and egg question and of course there is a kind of circularity: Eve must have had a mother and so forth and so on. So what I think is that bootstrapping is indeed the only option, and that is precisely how evolution is supposed to work. How can there ever be hammers? In order to make one you need a smith and a fire and an anvil and a hammer. Circularity! But of course, you can make better hammers with inferior ones, back to where you only have stones and hand axes.

Moreover, the very start of becoming the very first person must occur between the new-born and whoever is taking care of her – as is indeed the case with every person-to-be. This interaction already involves some shared affectivity in all mammals, I guess. But if these caretakers are wolves, probably not much of what we think is a person will ever evolve – see the empirical evidence on Oxana Malaya and Genie that I mention in my book.¹⁰ If these caretakers are one of the non-human primates, we would recognise such an individual much more as ‘one of us’, though language as we know it would still be missing and that is probably a huge obstacle for conferring personhood to her. Of course, our Eve would never come into contact with a group of present-day persons, but only with other primates that have grown up with primate caretakers. And of course, according to evolution: we all started as just another kind of primates. Personhood is, in that respect, not much different from language, I guess. You learn your language from a competent language-user, but there was a time when there just weren’t any language users. I guess that personhood was bootstrapped in roughly the same way that language was: from the first communicative cries to warn your conspecifics, language bootstrapped itself. And from the first reactive attitudes or moral sentiments of gratitude and resentment, the habit of holding some individuals responsible (but not others) bootstrapped itself into something like personhood.

So now for the future. I don’t know if your mention of robots is relevant here; I’ll ignore it. As I understand it, your question is about the two isolated groups that grow apart in their habits and attitudes regarding personhood. These attitudes are, I think necessarily, still grounded in shared affectivity and shared intentionality, and on reactive attitudes of gratitude and resentment. So far so good. But still their attitudes might be different in other respects. What we are talking about is some cultural difference in what the groups accept as

¹⁰ Meijssing (2018): 232-233.

candidate person. That is nothing fictional of course: the notions of personhood have changed over time, and they can grow apart in societies that are isolated from one another. And according to my own definition, you would indeed have to say that there are two different types of person: person₁ and person₂. But what happens if the two groups come into contact again?

Well, I think that if just one person from group₁ would come into contact with group₂, they should treat her as a *prima facie* candidate for personhood₂. The *Old Testament* is full of horrible genocides, and the killing of ‘all who piss against the wall’, i.e. males, but at the same time it repeatedly admonishes the people to treat the *sojourner*, the stranger in the land, as one of your own. The stranger is on her own, and should be treated as all the members of group₂. Of course, this will lead to misunderstandings and faulty expectations on both sides, but they can be resolved: I guess that the person₁ will quickly change into a person₂ – she would assimilate. It gets more difficult when whole groups of persons₁ come into group₂. Here we have the situation of mass immigration, and that can happen without a tendency to assimilation. The newcomers would say that their criteria for personhood now belong as much to the whole group as those of the original group. Yet persons₁ tend not to fully recognise persons₂ and vice versa. Here we have the all too common situation of discrimination. It happens. And it is wrong.

So how can I claim that in the first situation the persons₂ should treat the one person₁ as one of their own? And how can I claim that discrimination is wrong? For I also claim that there is no metaphysical category of persons. So the persons₂ cannot make a *metaphysical* mistake. But they can make an ethical one. I would like to base this on some moral realism, but I am really not at home in that discussion, and I gather there is a lot to be said against moral realism. Yet what I think would be a good start is some kind of capabilities approach. By treating human beings as non-persons, you exclude

them from many social activities and interactions, and you keep them from fully exercising their capabilities and that is morally wrong.

I guess the situation is somewhat like a society where women have to observe *purdah*, and are not be seen outside their own homes. They are kept apart by walls or curtains within the house, and have to wear extensive covering like the *burqa* if they are allowed out of the house at all. It is clear that within the women's quarters, the women are fully persons among themselves – like a group². Nevertheless, their capabilities are severely restricted, personally, socially and economically. Restricting capabilities is in itself morally wrong, I would say. Of course, in communities where *purdah* is practiced, people think it is morally wrong if women were to appear in public and take part of public life. But then what counts as morally right and wrong has to be negotiated all the time. In the same sense, I guess that what counts as a person has to be negotiated all the time. But whereas in the matter of personhood I don't think there is any further grounding of right and wrong, I have the intuition that in the matter of ethics there is a further grounding, and capabilities may be useful as an indication of that grounding.

That answer invites other questions, but that only seems proper for an interview with a philosopher! So, thank you so much Monica for your elaborate answers!

REFERENCES

- Asimov, I. (1978), *The bicentennial man and other stories*. New York, NY: Fawcett Crest.
- Goffman, E. (1956), *The presentation of self in everyday life*. Edinburgh, UK: University of Edinburgh Social Science Research Centre.
- Machiavelli, N. (1532/1976). *De Heerser (Il Principe)*. Amsterdam, The Netherlands: Athenaeum Polak & Van Gennepe
- McEwan, I. (2019). *Machines like me*. London: UK, Jonathan Cape.
- Meijnsing, M. (2018). *Waar was ik toen ik er niet was? Een filosofie van persoon en identiteit*. Nijmegen, The Netherlands: Vantilt.
- Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417-457.

Strawson, P. F. (1962/1982). Freedom and resentment. In G. Watson (Ed.) *Free will* (pp. 59-81). Oxford, UK: Oxford University Press.

TWEE KEER LEVENSLANG, TWEE KEER WETENSCHAPSFILOSOFISCHE FOUTEN

Ton Derksen

VOORWOORD

Monica Meijsing, wetenschapsfilosoof en psycholoog, heeft mij na mijn emeritaat zeer geholpen bij de analyse van de zaak-Lucia de B. Gelukkig was ze er wel, toen ze kritiek gaf.¹ Ook ten aanzien van andere zaken hebben ik van haar kritiek mogen profiteren. Daar ben ik haar zeer erkentelijk voor. Hieronder geef ik aan waarom juist een wetenschapsfilosoof bij strafrechtzaken een rol heeft te vervullen.

INLEIDING

Je zou denken: wat doen die wetenschapsfilosofen op het gebied van strafzaken? Wat weten ze nou van het strafrecht? De directe implicatie is correct: wetenschapsfilosofen zijn geen juristen. We hebben er dus niet voor gestudeerd. De indirecte *conversational implicature* (wetenschapsfilosofen hebben er niets te zoeken) is echter niet correct. Als wetenschapsfilosofen zijn we in feite beter geëquipeerd voor een strafzaak dan een jurist. Uiteindelijk gaat het daar om de vraag: Heeft Jan Klaassen Katrijn vermoord? En dat is primair een empirische vraag, geen juridische vraag. Als je het zo bekijkt, zijn juist de

¹ Zie, voor de doordenking van haar aanwezigheid, haar spannende studie Meijsing (2018). *Waar Was Ik Toen Ik Er Niet Was?*

officiëren van justitie, de rechters en helaas ook de advocaten niet opgeleid voor hun taak. Omdat ze geen wetenschappelijke scholing hebben, of preciezer omdat ze geen empirisch wetenschappelijke scholing hebben, en zich wel beroepsmatig met waarheidsvinding bezighouden, doen ze dat laatste op de manier zoals ze dat in de alledaagse praktijk geleerd hebben, met alle praktische voordelen maar ook geplaagd door alle alledaagse valkuilen in de waarheidsvinding. Ik bespreek hieronder twee zaken waarin ten onrechte een levenslang is uitgesproken: Olaf Hamers en Lucia de Berk. De fout kan met wetenschapsfilosofische middelen simpel aangegeven worden.²

COGNITIEVE INSTINCTEN ALS EERSTE LEIDRAAD BIJ WAARHEIDSVINDING

Vele filosofen hebben ons geleerd dat een mens een rationeel wezen is, en er goed aan doet rationeel te zijn. Het zou zo mooi zijn als dat kon, maar de tijd en het brein die ons ter beschikking staan, laten ons dat maar beperkt toe. Er zijn altijd weer mogelijkheden waaraan we niet gedacht hebben, of waarvoor we de tijd niet hadden om over na te denken. Toch gaat het in het alledaagse leven tamelijk goed met ons. We zijn in elk geval erin geslaagd om te overleven. Hoe kan dat, terwijl we niet de wijze raad van de Ouden kunnen opvolgen? Hier helpt de evolutie ons een handje. We zijn zó uitgeselecteerd dat onze eerste reacties ons een bovengemiddelde kans geven om te overleven. Zo is bijvoorbeeld ons perceptieapparaat behoorlijk betrouwbaar. We hebben echte tijgers en echte auto's doorgaans op tijd in de gaten. Zo hebben we ook een hele serie redelijk betrouwbare cognitieve instincten die onze eerste reactie bepalen, of beter die de eerste reactie van ons brein vormen. Deze reacties zijn niet primair op waarheid gericht, maar op overleven (en dus ook op snelheid en *safety first*). Maar omdat kennis van waarheid – en ik bedoel hier

² Voor de zaak-Lucia de B., zie: Derksen (2006). Voor de zaak-Olaf Hamers, zie Derksen (2013).

correspondentie-waarheid, de enige waarheid waaraan je bij het gevecht om te overleven wat hebt – je doorgaans betere overlevingskansen biedt, speelt waarheid indirect een rol als selectie criterium voor die cognitieve instincten.

Veel van onze oordelen zijn in feite waarschijnlijkheidsoordelen ten aanzien waarvan we iets te zeker zijn. Ook hier helpen cognitieve instincten ons die – in het alledaagse leven gecombineerd – hun waarde hebben bewezen maar individueel en in een andere context zich zo maar tegen ons kunnen keren.

Ik geef twee voorbeelden van cognitieve instincten waarmee we goed weggkomen: coherentie als waarheidscriterium (het bewijsmateriaal past bij het scenario, dus is het scenario waar) en het confirmatie-oordeel (we zijn geneigd te geloven wat bij ons standpunt past; wat niet past nemen we minder serieus). Het eerste houdt in dat we niet verder zoeken wanneer we passend bewijsmateriaal hebben gevonden. Dat is effectief en werkt vaak goed.

Het tweede, het confirmatieoordeel, staat tegenwoordig ook bekend als tunnelvisie. We weten allemaal dat dat heel slecht is, maar in de praktijk van alledag is het juist een zegen. Als ik mij elke keer dat Uri Geller op de TV komt, moet openstellen voor de mogelijkheid dat hij deze keer wel iets paranormaal kan presteren, had ik geen leven: ik heb hem ooit bestudeerd, nu is het wel genoeg. De man is een goochelaar. Hier komt tunnelvisie in feite neer op: geleerd hebben van de ervaring. Zolang dat achter de tunnelvisie zit, is tunnelvisie – als uitgangspunt – praktisch profijtelijk. Het spaart tijd en het werkt doorgaans.

VALKUILEN IN DE ALLEDAAGSE WAARHEIDSVINDING

Tot nu toe heb ik de voordelen van de cognitieve instincten geschetst. Maar omdat ze niet primair op waarheid zijn gericht, zitten er qua waarheidsvinding ook ernstige tekortkomingen aan die instincten, vooral wanneer ze buiten de

alledaagse context worden gebruikt, waar er geen ervaring is op basis waarvan ze opereren. En dat is precies het probleem in de strafrechtcontext. Ik noem hier een paar cognitieve instincten die, doordat hun werking niet onderkend werd, ten onrechte tot zware straffen hebben geleid.³

DE VERLEIDING VAN PASSEND BEWIJSMATERIAAL

Wanneer we voor ons standpunt steun vinden, verheugt ons dat. Vinden we nog meer ondersteunend bewijsmateriaal, dan achten we ons standpunt goed bevestigd, en hebben we na verloop van tijd nog steeds louter bevestigend bewijsmateriaal, dan stoppen we met het onderzoek. Dat is niet alleen wat wij in het alledaagse leven doen, ook de officier van justitie doet dat. Hij sluit tevreden zijn requisitoir af: ‘Alles past, meneer de rechter’. Hij beseft niet dat passend bewijsmateriaal nog niets zegt. Hoeveel je er ook van verzamelt, de vraag blijft altijd: Is dat bewijsmateriaal ook te rijmen met een alternatief scenario? Discrimineert het bewijsmateriaal tussen het eigen OM-scenario en een alternatief onschuld-scenario? Dat is de vraag. Olaf Hamers zit levenslang onder andere omdat de officier van justitie, en waarschijnlijk ook de rechter, getroffen werd door hoe goed zijn bewijsmateriaal bij het moordenaar-scenario paste.

DE ZAAK OLAF HAMERS

Volgens eigen zeggen was Olaf Hamers een tweedehands auto aan het kopen en even naar de WC gegaan, toen een man het kantoortje binnenkwam en de autohandelaar en zijn vrouw doodschoot. Hun kleindochter kwam er levend af, helaas wel met een schot door het hoofd. Van die moordenaar zijn geen sporen gevonden. Olaf rende weg met de sleutels van zijn pas gekochte auto, reed naar huis, haalde daar zijn Roemeense vriendin op en reed richting

³ Zie: Derksen (2010/2019).

Roemenië. Hij kwam na twee dagen terug. Olaf vertoonde bovendien ander gedrag dat in de ogen van het OM typisch gedrag van een dader is. Hij waarschuwde geen van de omstanders, hij nam geen contact op met de politie, hij belde ook niet om een ambulance, hij vluchtte meteen in de gestolen auto, en toen hij zag dat hij de dans niet kon ontspringen, kwam hij terug, maar hij nam wel even de tijd om een goed verhaal te bedenken. En toen – zo zag de officier het – meldde Olaf zich als de schuldbevuste dader, niet als getuige. Het is waar, al dit gedrag van Olaf is te duiden als gedrag van de dader. ‘Alles past’, om de afsluitende frase van de officier te gebruiken.

HET BELANG VAN EEN ALTERNATIEF SCENARIO

Maar al deze gegevens passen ook bij het scenario waarin Olaf Hamers onschuldig is. Hij had al betaald, het geld was weliswaar weg, maar de moordenaar kan dat meegenomen hebben, er lag in elk geval een betaalbon met de handtekening van de autohandelaar (OM: Olaf heeft eerst laten tekenen, en heeft pas daarna geschoten). Olaf vluchtte niet, hij moest zijn Roemeense vriendin vanwege een verlopen visum snel naar Roemenië terugrijden (tien getuigen bevestigen dat het plan al weken vastlag), hij nam geen contact met de politie op omdat ‘vrienden’ bij hem een wietplantage op zolder hadden (hij was vrachtwagenchauffeur, die weinig thuis was), hij heeft inderdaad geen ambulance gewaarschuwd. Sommigen mensen die ik spreek, vinden dit laatste een bewijs dat hij de moorden gepleegd moet hebben, maar psychologen wijzen op het gebrek aan bereidheid bij vele mensen om zich ergens mee te bemoeien. Rechercheurs op de politieacademie waar ik een college over deze zaak gaf, achtten dit niet-waarschuwen nietszeggend. Het bewijsmateriaal past bij het OM-scenario, maar evenzeer bij het onschuld-scenario. Dat alles past, heeft dus geen bewijswaarde.

DE ROL VAN NEGATIEVE ONDERZOEKSRESULTATEN

De politie probeert daderschap te bewijzen. Bijvoorbeeld, de politie heeft in de zaak Olaf Hamers uitgebreid de auto inclusief de kofferbak op schotresten laten onderzoeken. Als Olaf Hamers de moordenaar was geweest, had die auto onder de schotresten moeten zitten. Volgens het oordeel van het OM, dat door de rechters is overgenomen, heeft hij vier keer op drie mensen geschoten. Meteen daarna is hij met een auto die hij zojuist had gekocht (dan wel geroofd) naar Wenen gereden. Na twee dagen was hij terug. Zijn T-shirt dat hij tijdens het schieten gedragen had, had in de achterbak gelegen. Schotresten lossen niet op, ze verdwijnen alleen omdat ze ergens van afvallen of ergens vanaf geveegd worden. De achterbak met de ruwe vloerbedekking is dus een ideale plaats om schotresten te vinden – als ze er zijn. De politie vond niets. Had de politie iets gevonden, dan was dit bewijs van moordenaarschap. Nu ze niets vond, zei die uitslag volgens de politie en het OM ineens niets meer. Het argument was dat het ook wel voorkomt dat een schot maar weinig schotresten produceert.

Natuurlijk bewijst afwezigheid van schotresten niet dat Olaf onschuldig is, maar je kunt niet zomaar negatieve, ontlastende resultaten terzijde schuiven. Je moet kijken naar hoe waarschijnlijk (*likely*) die uitslag is uitgaande van de beide scenario's: (1) Olaf is de schutter, en (2) Olaf is niet de schutter. En ervan uitgaande dat hij niet de schutter is, is het veel waarschijnlijker (meer te verwachten, meer *likely*) dat geen schotresten worden gevonden dan ervan uitgaande dat hij de schutter is. En dat is gunstig voor Olaf. (De rol van *likelihoods* komt later ter sprake.)

Het alleen maar zoeken naar positief bewijsmateriaal is aldus een consequentie van ons natuurlijk cognitief instinct om met passend bewijsmateriaal genoeg te nemen. Het schaadt de onschuldige zeer.

IMMUNISEREN

De wetenschapsfilosoof Karl Popper heeft erop gewezen dat bepaalde uitspraken die op het eerste gezicht heel succesvol zijn omdat ze altijd waar lijken, juist loos zijn omdat ze geen enkele informatie bevatten. Het weerbericht ‘Misschien krijgen we morgen regen, maar misschien blijft het droog’ lijkt heel wat, maar het zegt in feite niets over het weer van morgen. Moet je nu wel je paraplu meenemen of niet? Het weerbericht is immuun voor falsificatie.

In de zaak Lucia de Berk immuniseert het Haagse hof frequent. Wanneer de advocaat erop wijst dat er nooit enig fysiek bewijs is gevonden voor de vermeende moorden van Lucia, terwijl ze op een kleine afdeling werkte met alle deuren altijd open en in het midden een balie met altijd iemand present, antwoordt het hof: Kun je na gaan hoe geraffineerd ze is. Wanneer Lucia bij de reanimatie helpt, is dat vreemd. Had ze de baby niet al willen doden? Maar nee, ze doet alsof. Wanneer ze niet helpt, toont dat haar onverschillig, koude karakter. Is ze van slag, dan is dat gespeeld. Is ze niet van slag, dan toont ze wederom haar gevoelloosheid. Zo is de verdachte altijd de klos. Niet omdat het bewijsmateriaal zo goed is, maar omdat het hof het zó construeert dat *elke uitslag* aantoont dat je de dader bent. Begrip van immuniseren had wellicht dit bewijsmateriaal ontzenuwd.

HAALBOOM: IMMUNISERINGEN DIE LEIDEN TOT 20 JAAR GEVANGENISSTRAF

Henk Haalboom is een grote fantast die bovendien zijn eigen intelligentie te hoog inschat. Toen hij van een moord werd verdacht op de jongere minnaar van zijn vriendin (feit: de jongere man had zijn vriendin *mogelijk* voor een kopje koffie uitgenodigd), wilde hij de politie helpen door er een mooi samenhangend verhaal van te maken. Dat liep uit de hand, de tegenspraken doken al snel op, en Haalboom werd in de ogen van de politie alleen maar

verdachter. Een stellige getuige die hem van foto's herkende en een psychologische/psychiatrische evaluatie als passief-agressief ('Meneer de rechter, elk moment kan de bom ontploffen!') leidden tot zijn veroordeling van twintig jaar. De vermeende herkenning is methodologisch een ramp (2^{de} keuze – mag niet – haar inschatting: de foto lijkt *voor 10%* op Haalboom; de rechter-commissaris herkende Haalboom niet van de foto), en de evaluatie als passief-agressief is een immunisering. Dit laatste leg ik verder uit.

Haalboom werd tijdens zijn voorarrest door een mannelijke psycholoog onderzocht. Hij vroeg Haalboom herhaaldelijk naar het waarom van de moord en citeerde daarbij uit het (belastende) politiedossier. Haalboom ergerde zich aan het vooringenomen onderzoek en liet dat duidelijk blijken. De psycholoog vatte dat in zijn evaluatie samen als: Haalboom is passief-agressief, dat wil zeggen: hij is agressief maar hij slaat er (nog) niet op los. Ook de vrouwelijke psychiater vroeg naar het waarom van de moord en ook zij citeerde daarbij uit het (belastende) politiedossier. Aan de vrouw toonde Haalboom niet zijn ongenoegen en reageerde op haar vraag. De psychiater concludeerde dat het vreemd was dat Haalboom niet agressief op haar kritische vragen reageerde. Dat moet agressie hebben opgewekt, maar hij bleek dat niet te uiten. Het gemeenschappelijk oordeel is: Haalboom is passief-agressief. Maar bedenk nu dat Haalboom *onafhankelijk van zijn reacties* als passief-agressief beoordeeld wordt. Of hij nu wel kwaad op de kritische bejegening reageerde of er niet op reageerde – wat Haalboom ook doet – de evaluatie is: passief-agressief.

Een psychiater legde mij uit dat dit voor een psychiater een gematigd oordeel is. Dat was het niet voor de officier van justitie, toen hij deze evaluatie aan de rechter uitlegde: Dit is een man die zo maar tot een agressieve daad in staat is. Hij lijkt vreedzaam, maar de agressie kan er elk moment uitkomen. Dat gebeurde toen hij door zijn jonge concurrent in de liefde werd gedwarsboomd. De stoppen sloegen door. De rechter was overtuigd.

Aldus kan een simpele immunisering en een methodologisch onverantwoorde ‘herkenning’ leiden tot twintig jaar gevangenisstraf.⁴

DE OMKERING

Ook het OM immuniseert er vaak lustig op los, en wel volgens de immuniserende strategie van de OMkering: het OM gaat uit van zijn gelijk – de verdachte is schuldig – en keert het argument simpelweg om, zonder enige onafhankelijke ondersteuning. Ik geef enkele voorbeelden uit de zaak van Olaf Hamers.

De verdediging argumenteert: ‘Olaf Hamers kon niet schieten. Dan lukt het je niet om drie mensen van een afstand elk met één schot door het hoofd te schieten’. *Het OM: Olaf had geluk.* (Hij moet geluk gehad hebben, want hij heeft vermoord en kon niet schieten.)

De verdediging: ‘Olaf was, zo zei iedereen, een vreedzaam type die juist ruzies vermeed’. *Het OM: Olaf was ten einde raad!* (Als hij een vreedzaam type is, moet hij wel ten einde raad geweest zijn, want hij heeft wel vermoord).

Verdediging: ‘Er lag een verkoopbon met zijn naam in de la van Jo. Een moordenaar afficheert zich niet op zo’n manier!’ *Het OM: Olaf was in paniek, hij heeft de bon laten liggen.* (Olaf was de moordenaar, hij heeft de bon laten liggen, dus hij moet in paniek zijn geweest).

Op deze manier is elk ontlastend argument te pareren. Het succes van het OM is gekocht door uit te gaan van het schuld-scenario, dat wil zeggen van zijn eigen gelijk.

⁴ Voor meer details zie: Derksen (2008).